The Relationship between Childhood Circumstances and Adult Health Disparities: Evidence from Colombia

Johanna Fajardo-Gonzalez
Department of Applied Economics, University of Minnesota
fajar016@umn.edu

Selected Paper prepared for presentation at the Agricultural & Applied Economics Association's 2014 AAEA Annual Meeting, Minneapolis, MN, July 27-29, 2014.

The Relationship between Childhood Circumstances and Adult Health Disparities: Evidence from Colombia

Preliminary version. Please do not cite.

This version: May, 2014

Abstract

This paper measures inequality of opportunity in adult health in Colombia employing the 2010 Living Standards and Social Mobility Survey. I study the relationship between childhood circumstances and health status in adulthood drawing on Roemer (1998)'s analytical approach to inequality of opportunity. I use stochastic dominance tests to capture differences in the conditional distributions of self-assessed health. This test is an initial assessment of inequality emerging from early life circumstances like parental education, household socioeconomic status in childhood and parental survival. I also calculate a dissimilarity index to provide a measure of inequality of opportunity in health and obtain the relative contributions of various circumstances using logistic regressions and the Shapley value decomposition. Since a limited set of circumstances are observed in the data, my estimation of inequality of opportunity provides a lower bound on the true inequality of opportunity. The findings suggest that there is substantial inequality of opportunity in adult health. Moreover, differences in household socioeconomic status during childhood and parental educational attainment appear to be the most important dimensions of inequality of opportunity in adult health.

Key Words: Health status, inequality, childhood circumstances, stochastic dominance, dissimilarity index, Shapley value, Colombia

1. Introduction

Evidence suggests that low socioeconomic status is related to poor health outcomes in both developed and developing countries. Within many countries, health outcomes appear to be worse in rural areas. Urban-rural differences could be partly explained by differences in access to and the quality of services and labor market conditions. This may also be the case for Colombia (MPS, 2010). Little is known, however, about the sources of adult health inequalities in urban and rural areas. I argue that exposure to different circumstances during childhood play an important role in the urban-rural differences in adult health outcomes currently observed.

I draw on Roemer (1998)'s theoretical approach to equality of opportunity to study the relationship between childhood circumstances and health status in adulthood. The concept of equality of opportunity was first proposed by Rawls (1971). Rawls emphasizes that personal responsibility is an important qualifier of the kind of equality that is ethically desirable and that opportunities should be equally open to all individuals regardless of factors that represent their origin or identity. In this line of research, Dworkin (1981) develops the concepts of equality of welfare and equality of resources, whereas Arneson (1989) and Cohen (1989) provide insights into equality of opportunity for welfare and equality of access to advantage. Sen (1980) has also contributed to this discussion with his economic theory of capabilities.

The inequality of opportunity approach developed by economist John Roemer (1998) is based on the view that the sources of an individual's desirable outcome, like good health or high income, can be separated between circumstances and responsibility factors (in Roemer, responsibility factors are referred to as "effort" variables) Circumstances are economically exogenous factors that cannot be controlled by an individual, and inequalities emerging from such circumstances should be compensated for. Conversely, responsibility factors are affected by individual choice and, therefore, inequalities arising from different individual responsibility are normatively acceptable. The most important implication is that an equal-opportunity policy would aim at providing everyone with the same opportunity to achieve or enjoy a good outcome. A social planner would seek to equalize opportunities rather than outcomes and would allow individuals to be fully responsible for their choices and final results. Reducing inequalities due to differences in circumstances is not only important from a social justice and fairness perspective. From a development standpoint, empirical evidence suggests that reducing inequality of opportunity also boosts human capital accumulation and growth (Marrero & Rodriguez, 2013)

I particularly build on the research on inequality of opportunity in health that has been recently put forward (Rosa-Dias, 2009; Trannoy et al., 2010; Li Donni, Peragine, & Pignataro, 2014). Rosa-Dias (2009) finds that about 21 percent of health inequality in adulthood, for a cohort of British individuals born in 1956, is mainly related to circumstances in childhood such as maternal education, spells of financial difficulties during childhood, as well as poor health and obesity in childhood. The empirical analysis developed in this paper is also grounded on Trannoy et al. (2010) and Li Donni, Peragine & Pignataro (2014). Trannoy et. al study inequality of opportunity among French adults and suggest that it would be halved if the effects of individual circumstances were removed. Li Donni, Peragine & Pignataro, in contrast to Rosa-Dias, apply an alternative empirical approach to data from various waves of the British Household Panel Survey and estimate that about 30 percent of adult health inequality is due to circumstances.

Trannoy et al. (2010) also discuss possible channels of health inequality transmission across generations. The first channel suggests that a specific risk that takes place during childhood has a direct influence on adult health following a latency period. The second channel suggests that human capital investments during childhood and the transmission of parental socioeconomic status have an indirect influence on health status in adulthood. The third channel assumes that health inequality is directly transmitted from one generation to the next. In this paper, I also provide suggestive evidence on these transmission channels.

This paper contributes to the literature by studying the differences in inequality of opportunity between residents in rural and urban areas. Health status varies greatly between rural and urban residents: 32 percent of the rural population in Colombia reports a poor or fair health status whereas 22 percent of the urban population reports a similar status. This urban-rural difference in health outcomes may be driven by different exposure to early life circumstances. I specifically address the following research question: among the set of observed circumstances, which particular early life circumstances have a salient long-term association with observed inequality of opportunity in rural and urban areas in Colombia? To the best of my knowledge this paper is among the first to answer these questions using data from a developing country.

I use data from the 2010 Colombian Living Standards and Social Mobility Survey. In the empirical analysis, I perform stochastic dominance tests to detect inequality of opportunity in the conditional distributions of self-assessed health status. In addition, I quantify the specific contributions of childhood circumstances such as parental education and household socioeconomic status at age 10 in observed health inequality through the decomposition of a dissimilarity inequality index using the Shapley value.

My findings, subject to some econometric caveats, raise an important implication. The socioeconomic background of an individual has a bearing on adult health status and policies should aim at reducing the intergenerational transmission of health inequalities. I identify in this paper some of the childhood circumstances that are most associated with adult health inequality. More importantly, this association differs for residents in urban and rural areas.

The rest of this paper proceeds as follows: Section 2 describes the 2010 Living Standards and Social Mobility Survey, and Section 3 explains the empirical methods. Finally, I present results in Section 4 and conclude in Section 5.

2. Data

The main source of data for this paper is the Colombian Living Standards and Social Mobility (LSSM) Survey carried out by the Colombian Bureau of Statistics (DANE) during 2010. This survey provides current and retrospective measures of socioeconomic characteristics. The LSSM is representative at the national level as well as for rural and urban areas. The survey also has national coverage by region[1]. The LSSM includes recall questions on living conditions when the respondent was 10 years old in a specific social mobility module. This set of questions provides information on parental educational attainment and ownership of some durable assets during childhood. The social mobility module in the LSSM only considers heads of household who are between 25 and 65 years old. The sample design ensures that the final sample of 2,253 individuals represent about 9.57 million heads of household in Colombia.

The outcome of interest is health status in adulthood. It is measured by self-assessed health status, which has been demonstrated to be effective in predicting mortality (Idler & Benyamini, 1997; van Doorslaer & Gerdtham, 2003) and health care utilization (De Salvo et al., 2005) Individuals rank their health as either poor (1), fair (2), good (3) or excellent (4) when answering the question “In general, how do you rate your health status?”. In this sample, around 73 percent of the respondents reported a good or an excellent health status whereas 2.2 percent reported a poor health status[2]. By subsample, 78 percent of urban residents reported at least a good health status whereas 68 percent of rural residents reported a similar status.

[1]Atlantic (Atlántico, Bolívar, Cesar, Córdoba, Guajira, Magdalena, Sucre), Eastern (Boyacá, Cundinamarca, Meta, Norte de Santander, Santander), Central (Caldas, Caquetá, Huila, Quindío, Risaralda, Tolima), Pacific (Cauca, Chocó, Nariño), Orinoquia-Amazonia (Arauca, Amazonas, Casanare, Guainía, Guaviare, Putumayo, Vaupés, Vichada), Antioquia, Valle del Cauca, San Andrés y Providencia, and Bogotá.

[2] See Tables 1a, 1b and 1c for a summary of descriptive statistics for the full sample, the urban subsample, and the rural subsample, respectively.

The set of *early-life circumstances* includes parental educational level, household socioeconomic status at age 10 and parental vital status. Parental educational attainment is a categorical variable that indicates whether a parent has completed or not a specific level of education (primary school, secondary school or higher) In this sample, approximately 60 percent of the heads of household reported that their parents did not attend school or did not complete primary education. In contrast, less than 9 percent indicated that their parents completed secondary school or a higher education level. In urban areas, 46 percent of fathers and 51 percent of mothers did not complete primary education. In rural areas, the percentages for incomplete primary education are higher than in urban areas: 54% for fathers and 62% for mothers.

Household socioeconomic status at age 10 is characterized by a categorical variable that indicates the socioeconomic status quintile a household falls into, based on an asset index computed using principal components analysis[3], following the methodology for wealth indices proposed by Vyas & Kumaranayake (2006). About 37 percent of the heads of household are assigned to the first quintile of the socioeconomic index, according to their reports of assets ownership[4]. The remaining individuals are evenly distributed across quintiles 2 to 5. In urban areas, the first (lowest) and fifth (highest) quintiles have approximately the same proportion of individuals: 25 percent. In rural areas, in contrast, 54 percent of individual belong in the first quintile.

Parental vital status is denoted by dichotomous variables that indicate whether a parent is alive at the time of the survey. In this sample, 48 percent of the household heads reported that their fathers were still alive at the time of the survey whereas only 65 percent reported having their mothers alive.

I also consider other variables that are likely to affect individual SAH. In the set of *other circumstances*, I also include ethnicity (indigenous, African Colombian, or none), urban or rural location of birth, and region of birth. In this sample, about 9 percent of household heads reported being a member of an ethnic minority (more specifically, 2.6 percent indicated indigenous roots and 6.4 percent indicated African Colombian roots). Indigenous minorities are mostly located in rural areas, in contrast with African Colombian minorities who are more evenly distributed between urban and rural areas. Regarding location of birth, most urban (rural) residents were born in urban (rural) areas.

The LSSM does not provide information on individual or parental health-related behaviors. The only circumstance that is partly an individual responsibility variable I can control for is years of education. Educational attainment has been shown to have a positive and large association with health (Lleras-Muney, 2005; Arendt, 2005; Cutler, Lleras-Muney & Vogl, 2008) The average number of years of education of the heads of household in this sample is of 7.02 years, that is, some years of secondary education.

Finally, additional controls include gender and age. In the full sample, about 71 percent of household heads are males. In rural areas this figure is of 79 percent whereas in urban areas 64 percent.

[Tables 1a, 1b and 1c about here]

3. Empirical Strategy

I use two different approaches to test for inequality of opportunity. The first is a non-parametric approach that relies on stochastic dominance comparisons of the conditional distributions of the health outcome. These distributions are conditional on subgroups (in Roemer (1998), a subgroup is referred to as "type")

[3] Variables in the socioeconomic status index include ownership of appliances like washing machine, vacuum cleaner, refrigerator, gas or electric stove, gas or electric oven, television set, availability of electricity, as well as ownership of dwelling, automobile, or motorcycle.

[4] One potential concern that arises from the use of these data is the recall nature of the early-life circumstances. A threat to my analysis comes from the possibility that the information reported is less accurate for longer recall intervals, in particular, for older adults regarding assets ownership in their childhood.

defined by the categories of a circumstance of interest. For instance, for a circumstance variable c with m categories I run $\frac{m!}{[(m-2)!]\ 2}$ pairwise stochastic dominance tests.

The stochastic dominance analysis, as most non-parametric methods, requires big sample sizes. Ideally, one would like to control for characteristics such as age and gender when comparing the distributions of health status conditional on different subgroups of circumstances. In practice, testing for stochastic dominance with smaller subsamples may cause the non-parametric test to be no longer useful (Trannoy et al., 2010) This is one of the disadvantages of the stochastic dominance approach. In consequence, I obtain point estimates for inequality of opportunity, controlling for age and gender, using a parametric approach. This approach relies on the econometric estimation of a non-linear model for health status. The predicted probability of reporting at least a good health status is used to calculate a dissimilarity index of inequality of opportunity that is decomposed using the Shapley value. The decomposition allows obtaining the contribution (in terms of correlation, not causation) of each circumstance to observed health inequality.

3.1 Stochastic Dominance and Inequality of Opportunity

Roemer (1998) defines equality of opportunity as a situation where individuals with similar responsibilities (or efforts) reach similar outcomes, regardless of their circumstances. More formally, under equality of opportunity, the probability distribution of health status H given responsibility e does not depend on circumstances C or C'. That is,

$$\forall\ C \neq C', \forall\ e,\ F(H|C,e) = F(H|C',e) \qquad [1]$$

where $F(H|C,e)$ denotes the cumulative probability function.

Lefranc, Trannoy & Pistolesi (2008) suggest that different health-related outcomes can be seen as alternative lotteries resulting from the effect of luck and other random factors that are equally distributed across individuals sharing the same responsibility and circumstances.[5] Therefore, a consistent definition of inequality of opportunity formulates that different conditional distributions of health can be ordered according to expected utility theory. In their paper, Lefranc, Trannoy & Pistolesi (2008) propose a criterion to assess inequality of opportunity using stochastic dominance relationships. The authors assume that health status is increasing in responsibility and that the relative responsibility can be inferred from the observation of health status and circumstances. Thus, inequality of opportunity is satisfied if and only if the distributions of health status conditional on different sets of circumstances can be ordered by first-order stochastic dominance, such that

$$\forall\ C \neq C',\ F(H|C) >_{FSD}\ F(H|C') \qquad [2]$$

As noted earlier, self-assessed health status is an ordinal variable. In this case, the stochastic dominance test is performed using a non-parametric test proposed by Yalonetzky (2013), as the more familiar statistical tests for stochastic dominance such as the Kolmogorov-Smirnov or the Davidson-Duclos cannot be directly applied to outcomes that are ordinal and lack any cardinal meaning.

The univariate extension of the test proposed by Yalonetzky[6] requires the introduction of further notation. Let A be the subgroup of individuals who share exposure to circumstance category a (e.g., individuals

[5] The authors also note that luck could lead to differences in individual health outcomes as long as it remains neutral with respect to circumstances.
[6] I heavily rely on Anand, Roope & Gray (2014) in the presentation of the test. The stochastic dominance test is formally introduced in Appendix 1.

whose mothers have incomplete primary education), and B the subgroup who share exposure to circumstance category b (e.g., individuals whose mothers have incomplete secondary education). The sample size of each group is denoted by n_A and n_B, respectively. Each individual in each group $g \in \{A, B\}$ reports a health status which lies in one of $S \in \mathbb{N}$ ordinal categories. Let $\boldsymbol{h}_g \epsilon \mathbb{N}_{\uparrow}^{N_g}$ be a vector of health status scores, where the $\uparrow$ subscript indicates that the ordinal categories are ordered in terms of their desirability, from the least to the most desired one. The *i-th* element of $\boldsymbol{h}_g$ is given by $h_{ig} \in \{1,2,3\}$.

For $k \in \{1,2,3\}$, $F_g(k) \equiv \Pr(h_{ig} \leq k)$ denotes the cumulative probability function. The difference in cumulative probability functions is defined as $\Delta F(\,) \equiv F_A(\,) - F_B(\,)$ The empirical estimates for the probability that a randomly selected individual from group g has a health status in category $j \in \{1,2\}$ are denoted by $\widehat{p_{JA}}$ and $\widehat{p_{JB}}$, respectively.

A test for the hypothesis that A does not first-order-stochastic dominate B against the alternative that A first-order-stochastic dominates B is given by

$$H_0 = \Delta F(k) \geq 0 \text{ for some } k \in \{1,2\} \qquad [3]$$

$$H_1 = \Delta F(k) < 0 \text{ for all } k \in \{1,2\}$$

The sample estimate for $\Delta F(k)$ is

$$\widehat{\Delta F}(k) = \sum_{j=1}^{k} \left(\widehat{p_{JA}} - \widehat{p_{JB}}\right) \qquad [4]$$

The corresponding z-statistic, z_k^l, is given by

$$z_k^l = \frac{\sum_{j=1}^{k}\left(\widehat{p_{JA}} - \widehat{p_{JB}}\right)}{\sqrt{\sum_{j=1}^{k}\left(\frac{\widehat{p_{JA}}\left(1-\widehat{p_{JA}}\right)}{n_A} + \frac{\widehat{p_{JB}}\left(1-\widehat{p_{JB}}\right)}{n_B} - \frac{\widehat{p_{JA}}}{n_A}\left(\sum_{l=1,l\neq j}^{k} \widehat{p_{lA}}\right) - \frac{\widehat{p_{JB}}}{n_B}\left(\sum_{l=1,l\neq j}^{k} \widehat{p_{lB}}\right)\right)}} \qquad [5]$$

H_0 is rejected if and only if $z_k^l \leq -z^* < 0$ for all $k \in \{1,2\}$, where $-z^*$ is the left-tail critical value from the standard normal distribution for a desired level of statistical significance.

I perform $\frac{m!}{[(m-2)!]\,2}$ pairwise tests for each circumstance variable *c* that has *m* response categories. To assess the differences in inequality of opportunity between urban and rural residents, I perform separate statistical tests for the sample of all individuals, the subsample of individuals residing in rural areas, and the subsample of individuals residing in urban areas.

3.2 The Dissimilarity Index of Inequality and its Decomposition through the Shapley Value

In the LSSM sample, 2.2 percent of the respondents report a poor health status (category 1) whereas 7.1 percent report an excellent health status (category 4). For the subsequent analysis, I group the two lower categories (1 and 2) and the two upper categories (3 and 4) to define a dichotomous variable which equals 0 if the respondent reports a poor or fair health status, and equals 1 if the respondent reports a good or excellent health status.

I measure inequality of opportunity using the dissimilarity inequality index, which has been used in inequality analysis using binary outcomes (Paes de Barros, Molinas & Saavedra, 2008; Paes de Barros et al., 2009). The dissimilarity index is a measure proportional to the absolute distance between the distribution of circumstances among those with high outcomes (i.e., excellent health) and the distribution among those with low outcomes (i.e., poor health). Paes de Barros, Molinas & Saavedra (2008) show that the dissimilarity index is also a measure of the absolute distance between the distribution of circumstances among those with high outcomes and the overall distribution of circumstances. Mathematically, Paes de Barros, Molinas & Saavedra (2008) show that the index is expressed as

$$D = \frac{1}{2}\sum_{k=1}^{M} |f_1(x_k) - f(x_k)| \quad [6]$$

where $f_1(x_k)$ denotes the distribution of circumstances among those who enjoy an excellent health and $f(x_k)$ the overall distribution of circumstances. M denotes the number of possible circumstances groups that can be formed with J circumstance variables c_j, with j=1,…,J. Each circumstance takes a value among g_j categories, then $M = \prod_{j=1}^{J} g_j$. The set of all possible values of the set of circumstances is $\{x_1, \dots, x_M\}$, with x_k a member of the set.

Paes de Barros, Molinas & Saavedra show that a consistent estimator for the dissimilarity index for binary outcomes is given by

$$\widehat{D} = \frac{1}{2\bar{p}}\sum_{i=1}^{n} w_i |\widehat{p_i} - \bar{p}| \quad [7]$$

where $\widehat{p_i}$ is the predicted probability of achieving a good or excellent health status for individual *i=1,…,n*. The estimated conditional probability is $\bar{p} = \sum_{i=1}^{n} w_i \widehat{p_i}$, where w_i denote sampling weights.

Gignoux & Ersado (2012) also show that the dissimilarity index for binary outcomes satisfies some important properties of inequality indices. The index equals 0 if the conditional distributions of health given circumstances are identical, and equals 1 when one individual always attains an excellent health while others do not. The index is insensitive to transfers of opportunities between circumstance groups that are above or below the average population achievement. Lastly, the index can only increase when new circumstances are added. Elaborating on the last property, Ferreira & Gignoux (2011) show that the measure of inequality of opportunity obtained with a set of observed circumstances is a lower bound on the true inequality of opportunity that would be captured if the full vector of circumstances was observed.

In order to know what circumstances correlate the most with observed inequality of opportunity I use the Shapley value decomposition. The Shapley value is a central solution concept in cooperative game theory that has been extended to inequality analysis by Shorrocks (2012). In this paper, I specifically follow Hoyos & Narayan (2012) and Gignoux & Ersado (2012) to perform this decomposition. The authors explain that the change in inequality that arises when a new circumstance is added to a set of circumstances depends on the sequence of inclusion of the different circumstance variables. The contribution of each circumstance is measured by the average change in inequality over all possible inclusion sequences. More formally, the change in the dissimilarity index when circumstance c is added to a subset M of circumstances is given by

$$\Delta D_c = \sum_{M \subset C \setminus \{c\}} \frac{|m|!(\kappa - |m| - 1)!}{\kappa!} [D(M \cup \{c\}) - D(M)] \quad [8]$$

where C denotes the entire set of κ circumstances, and M is a subset of C that includes m circumstance variables but c. $D(M)$ is the dissimilarity index for the subset M and $D(M \cup \{c\})$ is the index obtained after adding circumstance c to the subset M.

Let $D(\kappa)$ be the dissimilarity index for the set of κ circumstances. Therefore, the contribution (in terms of correlation, not causation) of circumstance κ to $D(\kappa)$ is defined by

$$S_c = \frac{\Delta D_c}{D(\kappa)} \text{ where } \sum_{i \in C} S_i = 1 \qquad [9]$$

As a result, I have an additive decomposition of the dissimilarity index that allows obtaining the contribution (in terms of correlation, not causation) of each circumstance to observed health inequality.

The dissimilarity index of inequality of opportunity can be interpreted as the minimum fraction of the number of healthier persons that need to be redistributed across circumstance groups in order to achieve equal opportunity, that is, an equal proportion of less healthy persons in all circumstance groups (Paes de Barros, Molinas & Saavedra, 2008)[7]. The index ranges from 0 to 1, with 0 indicating a situation with equality of opportunity.

Empirically, the calculation of the dissimilarity index first requires the estimation of a logistic regression model to obtain $\hat{p_i}$. In the following sub-section, I provide further details of the model to be estimated. Note that the results from this analysis do not provide any causal interpretation. In this study, I am mainly interested in providing a measurement of inequality of opportunity related to early life circumstances.

3.2.1 Logistic Regression Model for the Relationship between Health Status and Early Life Circumstances

In this sub-section I provide the empirical specification for the logistic regression model. The predicted probability of achieving a good or excellent health status is obtained after the estimation of the logit model. Thereafter, I calculate the dissimilarity index. This procedure is performed for the full sample, and for urban and rural areas subsamples.

First consider the health production function

$$\mathbf{H} = f(C, D, e, u) \qquad [10]$$

where C is a vector of individual circumstances, D a vector of demographic controls and e a vector of responsibility factors. The residual term u captures luck and other random factors that are not captured by the other variables in the health production function. Notice that the responsibility variables can also be affected by individual circumstances. For instance, previous studies suggest that an individual choice variable like educational attainment is associated with a circumstance like parental socioeconomic background since more educated parents provide more inputs into the production of education of their children (Hanushek, 1986) This implies that,

$$\mathbf{H} = f(C, D, e(D, C, v), u) \qquad [11]$$

We can empirically approximate this relationship using a linear specification. Let $\mathbf{H}_i$ be the health outcome for individual i, e_i the vector of individual responsibility variables, and u_i and v_i be error terms that capture luck and other random factors. Equation [10] may be estimated using the system of equations

$$\mathbf{H}_i = C_i\alpha + e_i\beta + D_i\vartheta + u_i \qquad [12]$$

[7] An alternative interpretation: the index indicates the percentage of available opportunities for enjoying a better health status that need to be reallocated from the adults who are healthier to the adults who are less healthy, in order to achieve equality of opportunity.

$$e_i = C_i\gamma + D_i\theta + v_i \quad [13]$$

For the purpose of this study, instead of estimating the full system of equations, I estimate a reduced-form using equations [12] and [13]. The reduced-form model for health status is given by

$$\mathbf{H}_{\boldsymbol{i}} = C_i\alpha + [C_i\gamma + D_i\theta + v_i]\beta + u_i$$

$$= C_i[\alpha + \beta\gamma] + D_i\theta\beta + v_i\beta + u_i$$

$$= C_i\varphi + D_i\psi + \varepsilon_i \quad [14]$$

where $\varphi = \alpha + \beta\gamma, \psi = \theta\beta$ and $\varepsilon_i = v_i\beta + u_i$

Thus, the parameter φ measures both the direct effect of circumstances and the indirect of circumstances through individual responsibility.

In this paper, the estimation of equation [14], using the LSSM data, only provides evidence of the correlation between early life circumstances and health status, and cannot be given any causal interpretation.

As mentioned in section 3, the following circumstances are observed in the 2010 LSSM data: ethnicity (E), father's highest educational level (FE), mother's highest educational level (ME), quintile of household socioeconomic status index during childhood (HS), urban or rural area of birth (LB), region of birth (RB), and parental health variables proxied by the dichotomous variables father alive in 2010 (FA) and mother alive in 2010 (MA). The only individual circumstance partly affected by individual choice that is observed in the dataset is years of education (ED). Demographic controls include gender (M) and age group (AG)

Roemer's definition of equality of opportunity assumes that efforts are orthogonal to circumstances. Hence, any other determinant of health status that is correlated with circumstances is also interpreted as a circumstance. For instance, a responsibility variable like educational attainment should be included in C_i given the part of it that is correlated with circumstances.

Health status is redefined as a binary variable given the characteristics of the LSSM data. Therefore, I study the mapping of equation [14] into the health status binary outcome H^*. More specifically, I estimate a logistic regression model for self-assessed health status controlling for circumstances, $C_i \equiv \{E_i, FE_i, ME_i, HS_i, LB_i, RB_i, FA_i, MA_i, ED_i\}$, and demographic controls, $D_i \equiv \{M_i, AG_i\}$

In order to estimate the global effect of observed circumstances on health status, I can also clean any variable included in C_i that is partly affected by individual responsibility of any influence coming from the other observed circumstances. In a recent study, Trannoy et al. (2010) proposed a two-step procedure to estimate the correlation of circumstances and health status in a non-linear model. The first step involves the estimation of the residuals from an auxiliary regression of the variable subject to individual responsibility on the full set of observed circumstances. In the second step, these residuals are included in the estimable health status equation along with the same vector of observed circumstances. Trannoy et al. emphasize that the residuals from step one represent individual responsibility and luck that allow an individual reaching a higher education level, for a given vector of observed circumstances. In this paper, I also adopt this empirical strategy.

In this case, the logistic regression model contains as an explanatory variable the term $\hat{v}_i^e$, which corresponds to the residuals obtained from the OLS estimation of the following model:

$$ED_i = k + C_i g + D_i w + v_i \qquad [15]$$

where v_i is a disturbance assumed to be normally distributed.

By construction, the residuals $\hat{v}_i^e$ are orthogonal to circumstances in the equation for health status and represent the share of individual educational attainment explained by individual responsibility, luck and unobserved characteristics, for the given vector of circumstances.

My interest is to gauge what circumstances are more correlated with the health status reported by residents in rural areas in contrast with that of respondents living in urban areas. I, therefore, estimate logistic regression models[8] for the sample of individuals residing in rural areas and the sample of individuals residing in urban areas using similar specifications to those presented in equations [15] to [17]. I do not perform this analysis using dichotomous variables for urban or rural residence because current residence is also considered a circumstance in Roemer's framework.

4. Results

4.1 Stochastic Dominance and Inequality of Opportunity

In this section, I empirically assess inequality of opportunity using the stochastic dominance approach. I analyze one circumstance at a time. In what follows, I refer to the group of individuals who share exposure to a particular circumstance category as "subgroup" (in Roemer (1998), a subgroup is referred to as "type").

In the LSSM data, health status is an ordinal variable which takes on values h=1, 2, 3, 4. I note in Section 3 that most responses concentrate in categories 2 (fair) and 3 (good). Thus, I group the lower two categories (1 and 2) to define a categorical variable which equals 1 if the respondent reports a poor or a fair health status, and equals 2 or 3 if the respondent reports a good or an excellent health status, respectively.

In this section, I particularly focus on the following *childhood circumstances*: parental education, household socioeconomic status at age 10, and parental vital status.

a. Parental Educational Attainment

To illustrate the application of the first-order stochastic dominance test in the context of the LSSM data, I define three subgroups based on maternal educational attainment: 1. Individuals whose mothers have incomplete primary school, 2. Mothers with complete primary school or incomplete secondary school, and 3. Mothers with complete secondary school or higher. Recall that higher values of the self-assessed health status denote a better health status reported.

I also define three subgroups based on paternal educational attainment, following the same definitions given for maternal educational attainment.

Figure 1 displays the cumulative probability distribution of health status conditional on mother's education level. Note that, for the full sample, the figure suggests that the distribution of health of individuals whose mothers have complete primary education or incomplete secondary (subgroup 2) dominates the distribution of health of individuals whose mothers have incomplete primary education or less (subgroup

[8] I retain both significant and insignificant coefficients in the estimation of the dissimilarity index, following Paes de Barros, Molinas & Saavedra (2008)

1). The sample of urban residents shows a similar pattern to that of the full sample. Regarding rural areas, the visual assessment is more difficult.

Figure 2 displays the distribution of health status conditional on father's education level. This figure also suggests that, for the full sample and for urban residents, the distribution of health for subgroup 3 dominates at first-order both the distribution for subgroup 2 and subgroup 1. In rural areas, the distribution of health conditional on complete primary may dominate the distribution of health conditional on incomplete primary.

I further examine the ranking of distributions using the non-parametric test proposed by Yalonetzky (2013). Table 2a displays the test results for the comparison of health status across different maternal education levels for all individuals in the sample. Comparing the distributions for the first two subgroups shown in Table 2a, panel a, at the 5% significance level and with a value of -z* of -1.645, the test suggests that the distribution for complete primary or incomplete secondary first-order-stochastically dominates the distribution for incomplete primary or no education in the LSSM sample. Regarding the first and the third subgroups (see Table 2a, panel b), the distribution for complete secondary or more dominates the distribution for primary education or less given the unanimously negative values of $\widehat{\Delta F}(k)$ and the significance of the z_k^l-statistic. A similar conclusion is suggested regarding the relationship between complete secondary or more and complete primary or incomplete secondary given the results presented in Table 2a, panel c. Given these results, I suggest that there is inequality of opportunity in adult health when a mother attains more education relative to a mother who obtains no more than some primary education.

[Tables 2a and 2b about here]

Regarding urban areas (see Table 2b), I find that the health distribution for mothers having completed secondary school dominate the health distribution for mothers who did not complete primary education. No dominance relationship can be established between the distribution for complete primary and incomplete primary as the z-statistic is not statistically significant for the first row, when I analyze the health category poor or fair. In rural areas (see also Table 2b), I find that no dominance relationship, at the first order, can be derived for the distributions of health status by each subgroup of maternal educational attainment.

The statistical test results for stochastic dominance using the subgroups defined by father's education level (see Table 3a) suggest that each of the distributions for complete primary and complete secondary dominates the distribution for incomplete primary at the first order. From my results, the dominance relationship between the distributions for complete primary and complete secondary is not clear. A similar result is obtained for the sample of urban residents, whereas no dominance relationship can be determined for rural residents (see Table 3b).

[Tables 3a and 3b about here]

b. Household Socioeconomic Status in Childhood

I define five subgroups using the quintiles of the socioeconomic status index calculated using information on ownership of assets by the individual's household at age 10. Regarding the sample of all individuals, from Figure 3, it is clear that the distribution for the fifth quintile dominates the distributions for the other four quintiles. The distribution for the first quintile is dominated by each of the distributions for the two upper quintiles at the first order of stochastic dominance. The non-parametric test results shown in Table 4a are in line with the graphical evidence about the order of these relationships. Similarly, they also suggest that the health distribution for the fourth quintile dominates the distribution for the second socioeconomic status quintile.

Turning to the urban subsample (see Table 4b), I find that the health distribution for the fifth quantile dominates each of the distributions for the first, second and fourth quintiles. These dominance relationships are statistically significant at the 5% level. My results also suggest that the distribution for the fourth quintile dominates the distribution for the second socioeconomic status quintile.
In contrast with the urban sample, it is hard to graphically define dominance relationships among the different distributions for the rural subsample (see Figure 3). The statistical tests results reported in Table 4c suggest that the only statistically significant dominance relationship is that of the health distribution for quintile 4 relative to quintile 1. From these results, I may argue that in Colombian rural areas, socioeconomic status during childhood may not be a highly influential circumstance in self-reported health status. Descendants do not necessarily report that they are healthier when their parents had a higher socioeconomic position. However, this result could also be driven by the recall nature of the question on assets ownership, in particular for older adults.

[Tables 4a, 4b and 4c about here]

c. Parental Vital Status

Figure 4 suggests that the distribution of health of individuals whose mothers are still alive dominates the distribution of health of individuals whose mothers are deceased, only for the full sample and the urban sample. A similar result is suggested for paternal vital status in Figure 5. The statistical test results shown in Table 5 suggest that the health distribution for mothers still alive dominates that of mothers deceased only for the full sample. I find that no dominance relationship, at the first order, can be derived for urban and rural areas. Similarly, the health distribution for fathers still alive dominates that of fathers deceased in both the full and the urban samples.

[Table 5 about here]

These results allow me to suggest that there is inequality of opportunity in health associated with parental vital status when the full sample is considered. In urban and rural samples, I am not able to draw a similar conclusion. Note that this analysis does not consider the effect of age. For instance, younger respondents may be more likely to report that their parents are still alive and also that their health status is at least good.

To close this section, the stochastic dominance analysis suggests that there is inequality of opportunity according to both parents' educational attainment, high parental socioeconomic status in childhood and both parents' vital status, when considering the full sample. For residents in urban areas, there is inequality of opportunity according to parents' education, high socioeconomic status, and paternal vital status. For residents in rural areas, I do not find conclusive evidence except for parental socioeconomic status in childhood. I cannot conclude whether there is inequality of opportunity in rural areas, according to the circumstances observed in the LSSM data.

4.2 Estimation Results from the Logistic Regression Model for Health Status

My interest is the calculation of the dissimilarity index. Section 3.2 notes that the index first requires the estimation of a logistic regression model since health status is defined as a binary outcome. In this section, I briefly describe the estimation results in order to suggest the potential direction of the correlation between reporting at least a good health status and the observed early life circumstances.

Results for the full sample

The first three columns in Table 6 display the estimation results of the logistic regression model for the full sample. In column 1, the results correspond to the estimation of the model controlling for years of education. In this sample, on average, males are potentially more likely to report a good health status than females. Similarly, there is a positive and significant correlation between reporting a good adult health status and high socioeconomic status during childhood (only in the comparison of quintile 3 against quintile 1, which is the excluded category). The estimated correlation between an individual's educational attainment, measured in years of education, and reporting a good adult health status is positive and highly significant.

Column 2 in Table 6 presents the results for the binary logistic regression model controlling for years of education purged from the effect of the other observed circumstances. First note that the OLS estimates of the equation for years of education are given in column 3 in the same table. As expected, a higher household socioeconomic status is positively correlated with obtaining more years of education. A positive association is also estimated for being born in urban areas relative to being born in rural areas. For the health status variable, note that, by construction, years of education purged from circumstances has the same point estimate and standard error as years of education. Note, however, that controlling for the correlation between years of education and the circumstance variables, does not change the direction of the basic relationships described in the previous paragraph, except for parental education. In particular, individuals whose fathers attained no more than some years of secondary education are also more likely to report a good health status, relative to those individuals whose fathers did not complete primary education. In the case of maternal education, the only significant and positive association with better health status is that of mothers having completed secondary education or more, relative to mothers with no education or some years of primary education. Finally, there is a positive and significant correlation between reporting a good adult health status and high socioeconomic status during childhood. Cleaning years of education from the influence of the observed circumstances allows obtaining significant and positive coefficient estimates for all quintiles of the socioeconomic status variables.

[Table 6 about here]

Results for rural and urban subsamples

Table 6 also presents the estimation results for the urban and rural samples. Regarding the results for the urban sample, I find that early life circumstances like household socioeconomic status and parental education have a significant effect on the likelihood of reporting at least a good health status. In particular, I find a positive relationship between reporting a good health status and coming from the third quintile of the socioeconomic status variable, having fathers who completed primary education and mothers who completed secondary education. This result also holds when I purge years of education from the influence of observed circumstances.

Using the sample for rural residents, I only find a positive and significant relationship between reporting a good health status and high socioeconomic status during childhood, only in the comparison of quintile 4 against quintile 1, which is the excluded category.

As a robustness check (results are not reported), I have also added variables for self-reported chronic illness and self-reported disability as control variables. Self-reported chronic is a dichotomous variable that indicates whether the individual suffers from a chronic or long-standing illness like diabetes, heart disease or cancer. Self-reported disability is also a dichotomous variable that indicates the presence of a permanent disability. These variables have a negative and significant effect on the likelihood of reporting a good health status. The relationship between circumstances and adult health status previously described do not change after including these health variables in my estimations.

Discussion on the potential transmission channels of health inequalities in adulthood

Trannoy et. al. (2010) discuss the possible channels of health inequality transmission across generations. The first channel suggests that a specific risk that takes place during childhood has a direct influence on adult health following a latency period. The second channel suggests that human capital investments during childhood and the transmission of parental socioeconomic status have an indirect influence on health status in adulthood. The third channel assumes that health is directly transmitted from one generation to the next. In this study, I provide suggestive evidence regarding the possible channels of transmission of health inequalities using the results of the logistic regression model.

For the full sample, I use the results in columns 2 and 3 in Table 6. Parental socioeconomic status and parental education attainment have both a direct effect and an indirect effect through education on self-reported health. These results may support the first and second transmission channels. Note that being born in urban areas has an indirect effect. Maternal vital status has no direct effect on health perceptions in adulthood as it would be expected according to the third channel of transmission.

The estimation results for the sample of urban residents also support the first and second channels according to parental socioeconomic status and parental education. In contrast, in rural areas the effect of parental socioeconomic status, parental education and maternal vital status is only indirect.

4.3 Dissimilarity Index of Inequality of Opportunity

I use the predicted probabilities from the estimation of the logistic regression models given by equations [14] and [15] to calculate the dissimilarity index. Table 8 displays the index value as well as its decomposition for the full sample, and for the rural and urban samples.

[Table 7 about here]

I begin with the analysis of the results for the full sample. The dissimilarity index obtained with the LSSM data is about 8.47%, when I include years of education in the vector of circumstances. The index is interpreted as the share of total opportunities for enjoying a better health status that would need to be redistributed from individuals who are healthier to individuals who are less healthy for equality of opportunity to prevail. The Shapley decomposition shows that the early life circumstances that have the largest contributions (in terms of correlation) to the dissimilarity index are household socioeconomic status at age 10 (15%), mother's education (10.5%) and father's education (10.4%). Once I clean years of education from the influence of circumstances, the decomposition of the index shows a slight increase in the contributions of socioeconomic status at age 10 (20%), mother's education (14%) and father's education (12%).

Turning to the results for the urban sample, I calculate a dissimilarity index of 8%, when I include years of education in the vector of circumstances. That is, 8% of total opportunities would need to be redistributed from individuals who are healthier to individuals who are less healthy for equality of opportunity to prevail. In rural areas, the index is relatively larger. About 10.1% of total opportunities would need to be redistributed from individuals who are healthier to individuals who are less healthy for equality of opportunity to prevail, when I include years of education in the vector of circumstances. The calculated indexes do not change considerably, once I clean years of education from the influence of circumstances. For urban areas, the decomposition of the index shows a slight increase in the contributions of socioeconomic status at age 10 (from 13% to 17%), mother's education (14% to 17%) and father's education (13% to 14%). For rural areas, the decomposition of the index shows a slight increase in the contributions of socioeconomic status at age 10 (from 19% to 25%) and paternal education (7% to 9%), the two circumstances that are most influential in inequality of opportunity in health status.

5. Concluding Remarks

This paper measures inequality of opportunity in health using the only dataset publicly available, the 2010 Living Standards and Social Mobility survey, that allows linking early life circumstances to current adult health conditions in Colombia. I have considered self-assessed health status as the outcome of interest as it is effective in predicting health care utilization and mortality. The early life circumstances include parental education, household socioeconomic status and an imperfect proxy for parental health. I also include other circumstances such as gender, ethnicity, educational attainment and place of birth that are likely to affect reported health status.

For the empirical analysis I use two different approaches. The first approach is stochastic dominance analysis. I assess inequality of opportunity by doing pairwise comparisons of the health distributions conditional on different subgroups of circumstances. I find that, for the full sample of individuals, there is inequality of opportunity due to parental educational attainment, parental socioeconomic status in childhood and parental vital status. In urban areas the pattern is very similar to that of the full sample. For residents in rural areas, I cannot conclude whether there is inequality of opportunity, according to the stochastic dominance tests using the circumstances observed in the LSSM data. Note that there is an interplay of circumstances that is not well captured by this analysis. Thus, in order to know what circumstances influence the most inequality of opportunity in adult health, I complement the stochastic dominance test with further analysis.

The second approach relies on the calculation of the dissimilarity index of inequality of opportunity. I use the Shapley value decomposition to obtain the relative contributions of early life circumstances. In line with previous studies that follow a similar approach, my findings suggest that household socioeconomic status during childhood is the circumstance that makes the largest contribution to inequality of opportunity. Parental educational attainment is highly associated with inequality of opportunity in health in urban areas but not so in rural areas. In contrast with urban areas, region of birth is potentially one of the most important circumstances in rural areas. These results are robust to the inclusion of two health variables that are likely to affect self-reported health status: permanent disability and chronic illness.

My study has several limitations. Self-reported health status may suffer from individual reporting heterogeneity. To the best of my knowledge, no study has provided evidence, appropriate for the Colombian context, in favor of or against the use of self-reported health in health research. Better measures of adult health status, like the body mass index, are not observed in the LSSM dataset. Unfortunately, more suitable surveys like the Demographic and Health Survey, do not provide intergenerational information. The study of inequality of opportunity in adult health in Colombia faces the usual problem of data availability.

An additional problem is the use of recall information. Household ownership of assets during childhood may not be accurately reported. This misreporting introduces bias in the estimates of the correlation between early life circumstances and adult health. Additionally, I would have preferred to use better parental health indicators than parental survival. My analysis does not allow to disentangle the effects of either genetic inheritance or parental health on investments in child's health capital, which is a weakness also identified in previous research (Trannoy et al., 2010)

The estimation of the dissimilarity index is also likely to be biased due to an omitted variables problem if any of the unobserved circumstances is correlated with any of the observed circumstances included in the analysis. Abras et al. (2013) showed that this problem is potentially mitigated since one of the properties of the dissimilarity index is that it can only increase when more circumstances are added. Of course, this property does not imply that the estimated contributions to the index also increase when more circumstances are included.

The inequality of opportunity analysis provides suggestive evidence of the lasting effects of childhood circumstances on adult health. The results presented in this study constitute a first step towards the identification of the potential channels through which health inequalities are transmitted from one generation to the next. They also suggest that education may play a key role in the reduction of health inequalities.

The transmission channels may also operate differently in rural and urban areas. In order to understand how these channels operate, further research might use exogenous variation coming from events like natural disasters or weather shocks. This analysis will require comprehensive information for at least two generations. Although a challenging research endeavor, this alternative may prove useful in disentangling the causal link between early life circumstances and adult health outcomes.

References

Abras, A., Hoyos, A., Narayan, A., & Tiwari, S. (2013). Inequality of opportunities in the labor market: evidence from life in transition surveys in Europe and Central Asia. *IZA Journal of Labor & Development*, 2(1), 7.

Anand, P., Roope, L., & Gray, A. (2013) Missing Dimensions in the Measurement of Wellbeing and Happiness. Mimeo.

Arendt, J. N. (2005). Does education cause better health? A Panel Data Analysis using School Reforms for Identification. *Economics of Education Review*, 24(2), 149-160.

Arneson, R. (1989). Equality of Opportunity for Welfare. *Philosophical Studies,* 56, 77–93

Cohen, G.A. (1989) On the Currency of Egalitarian Justice, *Ethics*, 99, 906–944

Cutler, D. M., Lleras-Muney, A., & Vogl, T. (2008). Socioeconomic Status and Health: Dimensions and Mechanisms, *National Bureau of Economic*, Research Working Paper No. 14333

DeSalvo,K., Fan, V.S., McDonnell, M.B., & Fihn, S.D. (2005) Predicting Mortality and Healthcare Utilization with a Single Question. *Health Services Research*, 40(4), 1234-46

Dworkin, R. (1981). What Is Equality? Part I: Equality of Welfare. *Philosophy & Public Affairs*, 3, 185-246

Fleurbaey, M. & Schokkaert, E. (2009). Unfair Inequalities in Health and Health Care. *Journal of Health Economics*, 28(1), 73-90

Formby, J., Smith, J., & Zheng, B. (2004). Mobility Measurement, Transition Matrices and Statistical Inference. *Journal of Econometrics*, 120, 181–205.

Gignoux, J. & Ersado, L. (2012) Labor-market entry and Opportunity Inequality in Egypt. Mimeo.

Gignoux, J., & Ferreira, F. (2011) The Measurement of Educational Inequality: Achievement and Opportunity, Policy Research Working Paper Series 5873, Washington, DC: The World Bank.

González, J. I., Sarmiento, A., Alonso, C. E., Angulo, R. & Espinosa, F.A. (2005) Efectos del nivel del ingreso y la desigualdad sobre la autopercepción en salud: análisis para el caso de Bogotá. In: *Revista Gerencia y políticas de salud*, 9:120-140. Bogotá: Universidad Javeriana (in Spanish)

Hanushek, E. A. (1986). The economics of schooling: Production and Efficiency in Public Schools. *Journal of Economics Literature,* 24(3), 1141-1177.

Howes, S. (1996) A New Test for Inferring Dominance from Sample Data. Discussion Paper, STICERD, London, UK: London School of Economics

Hoyos, A. & Narayan, A. (2011) Inequality of opportunities among children: how much does gender matter? Working Paper 76511. Washington DC: World Bank.

Idler, E., & Benyamini, Y. (1997) Self-rated Health and Mortality: A Review of Twenty-Seven Community Studies. *Journal of Health Economics*, 23(6), 1083-99

Jusot, F., Tubeuf, S., & Trannoy, A. (2010). Inequality of opportunities in health in Europe: Why so much difference across countries? Health, Econometrics and Data Group (HEDG) Working Papers 10/26.

Lefranc, A., Trannoy, A., & Pistolesi, N. (2008). Equality of opportunity: definitions and testable conditions with an application to income in France. THEMA Working Paper Series No. 2006-13, University of Cergy-Pontoise.

Lleras-Muney, A. (2005). The relationship between education and adult mortality in the United States. *The Review of Economic Studies*, 72(1), 189-221.

Li Donni, P., Checchi, V., & Pignataro, G. (2014), Ex-Ante and Ex-Post Measurement of Equality of Opportunity in Health: A Normative Decomposition, *Health Economics*, 23, 182-198

McDonough, P., Duncan, G.J., Williams, D., & House, J. (1997) Income dynamics and adult mortality in the United States, 1972 through 1989. *American Journal of Public Health*, 87(9), 1476-1483.

Marrero, G., & Rodriguez, J. (2013). Inequality of Opportunity and Growth. *Journal of Development Economics*, Volume 104,107-122

Ministerio de Protección Social (MPS) (2010) Análisis de las Desigualdades e Inequidades en Salud en Colombia, in *Análisis de la Situación de Salud en Colombia 2002-2007*, Tomo VI (in Spanish)

Paes de Barros, R., Molinas, J., & Saavedra, J. (2008). Measuring Inequality of Opportunities for Children. Washington, DC: World Bank.

Paes de Barros, R., Ferreira, F., Molinas, J., & Saavedra, J. (2009). Measuring Inequality of Opportunities in Latin America and the Caribbean. Washington, DC: The World Bank

Rawls, J. (1971) *A Theory of Justice*. Cambridge, MA: Harvard University Press

Roemer, J.E. (1998) *Equality of Opportunity*. Cambridge, MA: Harvard University Press

Rosa Dias, P. (2009). Inequality of opportunity in health: Evidence from a UK cohort study. *Health Economics*, 18(9), 1057–1074.

Sen, A. (1980) Equality of what? in S. McMurrin (ed.) The Tanner Lectures on Human Values, Salt Lake City: University of Utah Press

Shorrocks, A. F. (2012) "Decomposition procedures for distributional analysis: a unified framework based on the Shapley value", *Journal of Economic Inequality*, 10(1), 1-28.

Trannoy, A., Tubeuf, S., Jusot, F., & Devaux, M. (2010). Inequality of opportunities in health in France: A first pass. *Health Economics*, 19(8), 921–938.

van Doorslaer, E., & Gerdtham, U. G. (2003) Does Inequality in Self-assessed Health predict Inequality in Survival by Income?: Evidence from Swedish data. *Social Science & Medicine*, 57(9), 1621-29.

Vyas, S., & Kumaranayake, L. (2006) Constructing Socio-Economic Status Indices: How to Use Principal Components Analysis. *Health Policy Plan*, 21(6), 459-68

Yalonetzky, G. (2013) Stochastic Dominance with Ordinal Variables: Conditions and a Test. *Econometric Reviews*, 32(1), 126-163

Appendix 1. Stochastic Dominance Test for Ordinal Variables

Anand, Roope & Gray (2014) provide the univariate extension of the stochastic dominance test proposed by Yalonetzky (2013). In this appendix, I follow closely Anand, Roope & Gray's notation.

Suppose there are N_g individuals in group $g \in \{A, B\}$. Each individual indicates a health status which lies in one of $S \in \mathbb{N}$ ordinal categories. Let $\boldsymbol{h}_g \in \mathbb{N}_{\uparrow}^{N_g}$ be a vector of health status scores, where the ↑ subscript indicates that the ordinal categories are ordered in terms of their desirability from the least to the most desired one. The i-th element of $\boldsymbol{h}_g$ is given by $h_{ig} \in \{1, \ldots, S\}$.

For $k \in \{1, \ldots, S\}$, let $F_g(k) \equiv \Pr(h_{ig} \leq k)$ denote the cumulative probability function. Furthermore, the difference in cumulative probability functions is defined as $\Delta F() \equiv F_A() - F_B()$

Now, let p_{kg} be the probability that a randomly selected individual from $G = \{1, \ldots, N_g\}$ has a health status in category $k \in \{1, \ldots, S\}$, and $\boldsymbol{p}_g \in [0,1]^S$ be the corresponding vector of probabilities. The empirical estimate of p_{kg} from a random sample $n_g \leq N_g$ is given by

$$\widehat{p_{kg}} = \frac{1}{n_g} \sum_{i=1}^{n_g} I(k_i) \qquad \text{[A1]}$$

where $I(k_i)$ is an indicator function that equals 1 when $k_i = k$.

Let $\widehat{\boldsymbol{p}_{\boldsymbol{g}}}$ be the vector of empirical estimates of $\boldsymbol{p}_g$. Formby, Smith & Zheng (2004) show that the corresponding asymptotic result is given by

$$\sqrt{n_g}\left(\widehat{p_g} - p_g\right) \overset{d}{\rightarrow} N(\mathbf{0}, \Omega_g) \qquad \text{[A2]}$$

where Ω_g is a S-dimensional covariance matrix whose (k,l)-th element is equal to $p_{kg}(1 - p_{kg})$ whenever $k = l$, and $-p_{kg}p_{lg}$ whenever $k \neq l$.

Thus, under the null hypothesis that groups A and B are identically distributed, $\Omega_g = \Omega$ for any $g \in \{A, B\}$, so that

$$(\widehat{p_A} - \widehat{p_B}) \overset{d}{\rightarrow} N\left(\mathbf{0}, \frac{n_A + n_B}{n_A n_B} \Omega\right) \qquad \text{[A3]}$$

The empirical estimate of Ω_g has corresponding elements $\widehat{p_{kg}}(1 - \widehat{p_{kg}})$ whenever $k = l$, and $-\widehat{p_{kg}}\,\widehat{p_{lg}}$ whenever $k \neq l$.

Let $\widehat{\Delta \boldsymbol{F}}$ be the S-vector with k-th element given by $\widehat{\Delta F}(k) = \sum_{j=1}^{k}\left(\widehat{p_{JA}} - \widehat{p_{JB}}\right)$ and $\boldsymbol{L}$ be a S-dimensional lower triangular matrix of ones. Under the assumption that A and B are independent, the estimated covariance matrix of the empirical difference in cumulative probability functions is given by

$$var\left(\widehat{\Delta \boldsymbol{F}}\right) = \boldsymbol{L}\left(\frac{1}{n_A}\widehat{\Omega}_A + \frac{1}{n_B}\widehat{\Omega}_B\right)\boldsymbol{L}' \qquad \text{[A4]}$$

Thus, for each $k \in \{1, \ldots, S\}$, the corresponding z-statistic z_k^l is obtained by dividing $\widehat{\Delta F}(k)$ by its respective standard error, which is given by the squared root of the k-th diagonal element of $var\left(\widehat{\Delta \boldsymbol{F}}\right)$. More formally,

$$z_k^l = \frac{\sum_{j=1}^{k}(\widehat{p_{JA}}-\widehat{p_{JB}})}{\sqrt{\sum_{j=1}^{k}\left(\frac{\widehat{p_{JA}}(1-\widehat{p_{JA}})}{n_A}+\frac{\widehat{p_{JB}}(1-\widehat{p_{JB}})}{n_B}-\frac{\widehat{p_{JA}}}{n_A}\left(\sum_{l=1,l\neq j}^{k}\widehat{p_{lA}}\right)-\frac{\widehat{p_{JB}}}{n_B}\left(\sum_{l=1,l\neq j}^{k}\widehat{p_{lB}}\right)\right)}} \quad [A5]$$

The rejection rule proposed by Howes (1996) suggests that H_0 is rejected if and only if $z_k^l \leq -z^* < 0$ for all $k \in \{1, \dots, S-1\}$, where $-z^*$ is the left-tail critical value for a desired level of statistical significance.

Appendix 1 References

Anand, P., Roope, L., & Gray, A. (2013) Missing Dimensions in the Measurement of Wellbeing and Happiness. Unpublished.

Howes, S. (1996) A New Test for Inferring Dominance from Sample Data. Discussion Paper, STICERD, London, UK: London School of Economics

Yalonetzky, G. (2013) Stochastic Dominance with Ordinal Variables: Conditions and a Test. *Econometric Reviews*, 32(1), 126-163

Table 1a. Summary Statistics: Full Sample

Heads of Household between 25 and 65 years old. Total Number of Observations: 2,253

Variable	Observations	Mean or Proportion	Std. Dev.
Outcome			
Self-assessed Health Status	2,253	2.78	0.60
Poor	49	2.2%	0.15
Fair	556	24.7%	0.43
Good	1,487	66.0%	0.47
Excellent	161	7.1%	0.26
Early-life Circumstances			
Household Socioeconomic Status at Age 10			
Quantile 1	841	37.3%	0.48
Quantile 2	347	15.4%	0.36
Quantile 3	262	11.6%	0.32
Quantile 4	383	17.0%	0.38
Quantile 5	382	17.0%	0.38
No Information on Assets	38	1.7%	0.13
Education Level of Father			
None or Incomplete Primary	1,258	55.8%	0.50
Complete Primary and Incomplete Secondary	377	16.7%	0.37
Complete Secondary or More	194	8.6%	0.28
Unknown Father's Education	422	18.7%	0.39
No Information on Father's Education	2	0.1%	0.03
Education Level of Mother			
None or Incomplete Primary	1,345	59.7%	0.49
Complete Primary and Incomplete Secondary	447	19.8%	0.40
Complete Secondary or More	171	7.6%	0.26
Unknown Mother's Education	288	12.8%	0.33
No Information on Mother's Education	2	0.1%	0.03
Other circumstances			
Ethnicity			
Indigenous	59	2.6%	0.16
Black, Mulato, Raizal or Palenquero	144	6.4%	0.24
No ethnic minority	2,050	91.0%	0.29
Years of Education	2,253	7.02	4.65
Born in Urban Area	1,103	49.0%	0.50
Born in Rural Area	1,144	50.8%	0.50
No Information on Area of Birth	6	0.3%	0.05
Parental Health Status			
Father Deceased	1,166	51.8%	0.50
Father Alive	1,073	47.6%	0.50
No Information on Father's Vital Status	14	0.6%	0.08
Mother Deceased	767	34.0%	0.47
Mother Alive	1,472	65.3%	0.48
No Information on Mother's Vital Status	14	0.6%	0.08
Additional Controls			
Male	1,598	70.9%	0.45
Age	2,253	44.77	11.01
Age group			
25-35	504	22.4%	0.42
35-45	594	26.4%	0.44
45-55	646	28.7%	0.45
55-65	509	22.6%	0.42

Source: 2010 Colombian LSSM Survey

Table 1b. Summary Statistics: Urban Subsample

Urban Subsample. Total Number of Observations: 1,263

Variable	Observations	Mean or Proportion	Std. Dev.
Outcome			
Self-assessed Health Status	1,263	2.85	0.60
Poor	25	2.0%	0.14
Fair	258	20.4%	0.40
Good	856	67.8%	0.47
Excellent	124	9.8%	0.30
Early-life Circumstances			
Household Socioeconomic Status at Age 10			
Quantile 1	308	24.4%	0.43
Quantile 2	198	15.7%	0.36
Quantile 3	149	11.8%	0.32
Quantile 4	268	21.2%	0.41
Quantile 5	327	25.9%	0.44
No information on assets available	13	1.0%	0.10
Education Level of Father			
None or Incomplete Primary	585	46.3%	0.50
Complete Primary and Incomplete Secondary	289	22.9%	0.42
Complete Secondary or More	177	14.0%	0.35
Unknown Father's Education	210	16.6%	0.37
No information on father's education	2	0.2%	0.04
Education Level of Mother			
None or Incomplete Primary	647	51.2%	0.50
Complete Primary and Incomplete Secondary	333	26.4%	0.44
Complete Secondary or More	151	12.0%	0.32
Unknown Mother's Education	130	10.3%	0.30
No information on mother's education	2	0.2%	0.04
Other circumstances			
Ethnicity			
Indigenous	22	1.7%	0.13
Black, Mulato, Raizal or Palenquero	80	6.3%	0.24
No ethnic minority	1,161	91.9%	0.27
Years of Education	2,253	8.83	4.54
Born in Urban Area	899	71.2%	0.45
Born in Rural Area	359	28.4%	0.45
No information on area of birth	5	0.4%	0.06
Parental Health Status			
Father Deceased	656	51.9%	0.50
Father Alive	601	47.6%	0.50
No information on father's vital status	6	0.5%	0.07
Mother Deceased	426	33.7%	0.47
Mother Alive	831	65.8%	0.47
No information on mother's vital status	6	0.5%	0.07
Additional Controls			
Male	811	64.2%	0.48
Age	2,253	45.13	10.96
Age group			
25-35	275	21.8%	0.41
35-45	315	24.9%	0.43
45-55	385	30.5%	0.46
55-65	288	22.8%	0.42

Source: 2010 Colombian LSSM Survey

Table 1c. Summary Statistics: Rural Subsample

Rural Subsample. Total Number of Observations: 990

Variable	Observations	Mean or Proportion	Std. Dev.
Outcome			
Self-assessed Health Status	990	2.69	0.58
Poor	24	2.4%	0.15
Fair	298	30.1%	0.46
Good	631	63.7%	0.48
Excellent	37	3.7%	0.19
Early-life Circumstances			
Household Socioeconomic Status at Age 10			
Quantile 1	537	54.2%	0.50
Quantile 2	142	14.3%	0.35
Quantile 3	106	10.7%	0.31
Quantile 4	128	12.9%	0.34
Quantile 5	51	5.2%	0.22
No information on assets available	26	2.6%	0.16
Education Level of Father			
None or Incomplete Primary	673	68.0%	0.47
Complete Primary and Incomplete Secondary	88	8.9%	0.28
Complete Secondary or More	17	1.7%	0.13
Unknown Father's Education	212	21.4%	0.41
Education Level of Mother			
None or Incomplete Primary	698	70.5%	0.46
Complete Primary and Incomplete Secondary	114	11.5%	0.32
Complete Secondary or More	20	2.0%	0.14
Unknown Mother's Education	158	16.0%	0.37
Parental Health Status			
Father Deceased	513	51.8%	0.50
Father Alive	470	47.5%	0.50
No information on father's vital status	7	0.7%	0.08
Mother Deceased	355	35.9%	0.48
Mother Alive	628	63.4%	0.48
No information on mother's vital status	7	0.7%	0.08
Other circumstances			
Ethnicity			
Indigenous	37	3.7%	0.19
Black, Mulato, Raizal or Palenquero	64	6.5%	0.25
No ethnic minority	889	89.8%	0.30
Years of Education	2,253	4.71	3.66
Born in Urban Area	204	20.6%	0.41
Born in Rural Area	785	79.3%	0.40
No information on area of birth	1	0.1%	0.03
Additional Controls			
Male	787	79.5%	0.40
Age	2,253	44.31	11.06
Age group			
25-35	229	23.1%	0.42
35-45	279	28.2%	0.45
45-55	261	26.4%	0.44
55-65	221	22.3%	0.42

Source: 2010 Colombian LSSM Survey

Table 2a. Distribution of Health Status by Mother's Education Level: Full Sample

a. Complete Primary vs. Incomplete Primary

Health Status	Incomplete primary school or none			Complete primary school or incomplete secondary school			$\widehat{\Delta F}(k)$	z_k^l
	Freq.	%	Cumul. %	Freq.	%	Cumul. %		
1 = Poor/Fair	379	28.17	28.17	76	16.99	16.99	-0.112	-5.179 ***
2 = Good	884	65.72	93.89	330	73.84	90.83	-0.031	-2.022 ***
3 = Excellent	82	6.11	100	41	9.17	100		
Total	1,345	100		447	100			

b. Complete Secondary y vs. Incomplete Primary

Health Status	Incomplete primary school or none			Complete secondary school or higher			$\widehat{\Delta F}(k)$	z_k^l
	Freq.	%	Cumul. %	Freq.	%	Cumul. %		
1 = Poor/Fair	379	28.17	28.17	14	8.02	8.02	-0.202	-8.354 ***
2 = Good	884	65.72	93.89	113	65.79	73.81	-0.201	-5.863 ***
3 = Excellent	82	6.11	100	45	26.2	100.01		
Total	1,345	100		171	100.01			

c. Complete Secondary vs. Complete Primary

Health Status	Complete primary school or incomplete secondary school			Complete secondary school or higher			$\widehat{\Delta F}(k)$	z_k^l
	Freq.	%	Cumul. %	Freq.	%	Cumul. %		
1 = Poor/Fair	76	16.99	16.99	14	8.02	8.02	-0.090	-3.282 ***
2 = Good	330	73.84	90.83	113	65.79	73.81	-0.170	-4.690 ***
3 = Excellent	41	9.17	100	45	26.2	100.01		
Total	447	100		171	100.01			

*** denote that the statistic is significant at the 5% significant level
Source: 2010 Colombian LSSM Survey

Note: The null hypothesis is given by $H_0 = \Delta F(k) \geq 0$ for some $k \in \{1,2\}$ and the alternative is given by $H_1 = \Delta F(k) < 0$ for all $k \in \{1,2\}$. $\widehat{\Delta F}(k)$ indicates the estimated difference between the cumulative probability functions, $\widehat{F_B(k)}$–$\widehat{F_A(k)}$, where $\widehat{F_B(k)}$ indicates the cumulative probability function for the subgroup in the most-right panel and $\widehat{F_A(k)}$ for the most-left panel, for row k. $\mathrm{H_0}$ is rejected if and only if $\mathrm{z_k^l} \leq -\mathrm{z}^* < 0$ for all $\mathrm{k} \in \{1,2\}$, where $-\mathrm{z}^* = -1.645$ is the left-tail critical value at the 5% significance level. No ordering can be established if the two values for z_k^l do not have the same direction.

Table 2b. Distribution of Health Status by Mother's Education Level: Urban and Rural Subsamples

Residents in Urban Areas

a. Complete Primary vs. Incomplete Primary

Health Status	Incomplete primary school or none			Complete primary school or			$\widehat{\Delta F}(k)$	z_k^l
	Freq.	%	Cumul. %	Freq.	%	Cumul. %		
1 = Poor/Fair	174	26.91	26.91	53	15.96	15.96	-0.110	-4.119 ***
2 = Good	426	65.88	92.79	248	74.54	90.5	-0.023	-1.204
3 = Excellent	47	7.22	100.01	32	9.5	100		
Total	647	100.01		333	100			

b. Complete Secondary vs. Incomplete Primary

Health Status	Incomplete primary school or none			Complete secondary school or higher			$\widehat{\Delta F}(k)$	z_k^l
	Freq.	%	Cumul. %	Freq.	%	Cumul. %		
1 = Poor/Fair	174	26.91	26.91	11	7.05	7.05	-0.199	-7.311 ***
2 = Good	426	65.88	92.79	99	65.65	72.7	-0.201	-5.336 ***
3 = Excellent	47	7.22	100.01	41	27.3	100		
Total	647	100.01		151	100			

c. Complete Secondary vs. Complete Primary

Health Status	Complete primary school or incomplete secondary school			Complete secondary school or higher			$\widehat{\Delta F}(k)$	z_k^l
	Freq.	%	Cumul. %	Freq.	%	Cumul. %		
1 = Poor/Fair	53	15.96	15.96	11	7.05	7.05	-0.089	-3.080 ***
2 = Good	248	74.54	90.5	99	65.65	72.7	-0.178	-4.489 ***
3 = Excellent	32	9.5	100	41	27.3	100		
Total	333	100		151	100			

Residents in Rural Areas

a. Complete Primary vs. Incomplete Primary

Health Status	Incomplete primary school or none			Complete primary school or			$\widehat{\Delta F}(k)$	z_k^l
	Freq.	%	Cumul. %	Freq.	%	Cumul. %		
1 = Poor/Fair	174	26.91	26.91	53	15.96	15.96	-0.110	-4.119
2 = Good	426	65.88	92.79	248	74.54	90.5	-0.023	-1.204
3 = Excellent	47	7.22	100.01	32	9.5	100		
Total	647	100.01		333	100			

b. Complete Secondary vs. Incomplete Primary

Health Status	Incomplete primary school or none			Complete secondary school or higher			$\widehat{\Delta F}(k)$	z_k^l
	Freq.	%	Cumul. %	Freq.	%	Cumul. %		
1 = Poor/Fair	174	26.91	26.91	11	7.05	7.05	-0.199	-7.311
2 = Good	426	65.88	92.79	99	65.65	72.7	-0.201	-5.336
3 = Excellent	47	7.22	100.01	41	27.3	100		
Total	647	100.01		151	100			

c. Complete Secondary vs. Complete Primary

Health Status	Complete primary school or incomplete secondary school			Complete secondary school or higher			$\widehat{\Delta F}(k)$	z_k^l
	Freq.	%	Cumul. %	Freq.	%	Cumul. %		
1 = Poor/Fair	53	15.96	15.96	11	7.05	7.05	-0.089	-3.080
2 = Good	248	74.54	90.5	99	65.65	72.7	-0.178	-4.489
3 = Excellent	32	9.5	100	41	27.3	100		
Total	333	100		151	100			

*** denote that the statistic is significant at the 5% significant level

Source: 2010 Colombian LSSM Survey

Note: The null hypothesis is given by $H_0 = \Delta F(k) \geq 0$ for some $k \in \{1,2\}$ and the alternative is given by $H_1 = \Delta F(k) < 0$ for all $k \in \{1,2\}$.$\widehat{\Delta F}(k)$ indicates the estimated difference between the cumulative probability functions, $\widehat{F_B(k)}$–$\widehat{F_A(k)}$, where $\widehat{F_B(k)}$ indicates the cumulative probability function for the subgroup in the most-right panel and $\widehat{F_A(k)}$ for the most-left panel, for row k. H_0 is rejected if and only if $z_k^l \leq -z^* < 0$ for all $k \in \{1,2\}$, where $-z^* = -1.645$ is the left-tail critical value at the 5% significance level. No ordering can be established if the two values for z_k^l do not have the same direction.

Table 3a. Distribution of Health Status by Father's Education Level': Full Sample

a. Complete Primary vs. Incomplete Primary

Health Status	Incomplete primary school or none			Complete primary school or incomplete secondary school			$\widehat{\Delta F}(k)$	z_k^l
	Freq.	%	Cumul. %	Freq.	%	Cumul. %		
1 = Poor/Fair	359	28.54	28.54	58	15.32	15.32	-0.132	-5.876 ***
2 = Good	829	65.87	94.41	281	74.52	89.84	-0.046	-2.711 ***
3 = Excellent	70	5.59	100	38	10.16	100		
Total	1,258	100		377	100			

b. Complete Secondary y vs. Incomplete Primary

Health Status	Incomplete primary school or none			Complete secondary school or higher			$\widehat{\Delta F}(k)$	z_k^l
	Freq.	%	Cumul. %	Freq.	%	Cumul. %		
1 = Poor/Fair	359	28.54	28.54	22	11.23	11.23	-0.173	-6.658 ***
2 = Good	829	65.87	94.41	123	63.23	74.46	-0.200	-6.240 ***
3 = Excellent	70	5.59	100	50	25.54	100		
Total	1,258	100		194	100			

c. Complete Secondary vs. Complete Primary

Health Status	Complete primary school or incomplete secondary school			Complete secondary school or higher			$\widehat{\Delta F}(k)$	z_k^l
	Freq.	%	Cumul. %	Freq.	%	Cumul. %		
1 = Poor/Fair	58	15.32	15.32	22	11.23	11.23	-0.041	-1.396
2 = Good	281	74.52	89.84	123	63.23	74.46	-0.154	-4.399 ***
3 = Excellent	38	10.16	100	50	25.54	100		
Total	377	100		194	100			

*** denote that the statistic is significant at the 5% significant level

Source: 2010 Colombian LSSM Survey

Note: The null hypothesis is given by $H_0 = \Delta F(k) \geq 0$ for some $k \in \{1,2\}$ and the alternative is given by $H_1 = \Delta F(k) < 0$ for all $k \in \{1,2\}$. $\widehat{\Delta F}(k)$ indicates the estimated difference between the cumulative probability functions, $\widehat{F_B(k)}$–$\widehat{F_A(k)}$, where $\widehat{F_B(k)}$ indicates the cumulative probability function for the subgroup in the most-right panel and $\widehat{F_A(k)}$ for the most-left panel, for row k. H_0 is rejected if and only if $z_k^l \leq -z^* < 0$ for all $k \in \{1,2\}$, where $-z^* = -1.645$ is the left-tail critical value at the 5% significance level. No ordering can be established if the two values for z_k^l do not have the same direction.

Table 3b. Distribution of Health Status by Father's Education Level: Urban and Rural Subsamples

Residents in Urban Areas

a. Complete Primary vs. Incomplete Primary

Health Status	Incomplete primary school or none			Complete primary school or incomplete			$\widehat{\Delta F}(k)$	z_k^l
	Freq.	%	Cumul. %	Freq.	%	Cumul. %		
1 = Poor/Fair	160	27.39	27.39	42	14.45	14.45	-0.129	-4.670 ***
2 = Good	387	66.11	93.5	216	74.88	89.33	-0.042	-2.002 ***
3 = Excellent	38	6.5	100	31	10.67	100		
Total	585	100		289	100			

b. Complete Secondary vs. Incomplete Primary

Health Status	Incomplete primary school or none			Complete secondary school or higher			$\widehat{\Delta F}(k)$	z_k^l
	Freq.	%	Cumul. %	Freq.	%	Cumul. %		
1 = Poor/Fair	160	27.39	27.39	20	11.14	11.14	-0.163	-5.419 ***
2 = Good	387	66.11	93.5	111	62.65	73.79	-0.197	-5.698 ***
3 = Excellent	38	6.5	100	46	26.21	100		
Total	585	100		177	100			

c. Complete Secondary vs. Complete Primary

Health Status	Complete primary school or incomplete			Complete secondary school or higher			$\widehat{\Delta F}(k)$	z_k^l
	Freq.	%	Cumul. %	Freq.	%	Cumul. %		
1 = Poor/Fair	42	14.45	14.45	20	11.14	11.14	-0.033	-1.054
2 = Good	216	74.88	89.33	111	62.65	73.79	-0.155	-4.120 ***
3 = Excellent	31	10.67	100	46	26.21	100		
Total	289	100		177	100			

Residents in Rural Areas

a. Complete Primary vs. Incomplete Primary

Health Status	Incomplete primary school or none			Complete primary school or incomplete			$\widehat{\Delta F}(k)$	z_k^l
	Freq.	%	Cumul. %	Freq.	%	Cumul. %		
1 = Poor/Fair	212	31.44	31.44	21	23.78	23.78	-0.077	-1.570
2 = Good	439	65.26	96.7	62	71.02	94.8	-0.019	-0.771
3 = Excellent	22	3.3	100	5	5.21	100.01		
Total	673	100		88	100.01			

b. Complete Secondary vs. Incomplete Primary

Health Status	Incomplete primary school or none			Complete secondary school or higher			$\widehat{\Delta F}(k)$	z_k^l
	Freq.	%	Cumul. %	Freq.	%	Cumul. %		
1 = Poor/Fair	212	31.44	31.44	2	13.57	13.57	-0.179	-2.103 ***
2 = Good	439	65.26	96.7	14	79.46	93.03	-0.037	-0.591
3 = Excellent	22	3.3	100	1	6.97	100		
Total	673	100		17	100			

c. Complete Secondary vs. Complete Primary

Health Status	Complete primary school or incomplete			Complete secondary school or higher			$\widehat{\Delta F}(k)$	z_k^l
	Freq.	%	Cumul. %	Freq.	%	Cumul. %		
1 = Poor/Fair	21	23.78	23.78	2	13.57	13.57	-0.102	-1.079
2 = Good	62	71.02	94.8	14	79.46	93.03	-0.018	-0.268
3 = Excellent	5	5.21	100.01	1	6.97	100		
Total	88	100.01		17	100			

*** denote that the statistic is significant at the 5% significant level
Source: 2010 Colombian LSSM Survey

Note: The null hypothesis is given by $H_0 = \Delta F(k) \geq 0$ for some $k \in \{1,2\}$ and the alternative is given by $H_1 = \Delta F(k) < 0$ for all $k \in \{1,2\}$. $\widehat{\Delta F}(k)$ indicates the estimated difference between the cumulative probability functions, $\widehat{F_B(k)}$–$\widehat{F_A(k)}$, where $\widehat{F_B(k)}$ indicates the cumulative probability function for the subgroup in the most-right panel and $\widehat{F_A(k)}$ for the most-left panel, for row k. H_0 is rejected if and only if $z_k^l \leq -z^* < 0$ for all $k \in \{1,2\}$, where $-z^* = -1.645$ is the left-tail critical value at the 5% significance level. No ordering can be established if the two values for z_k^l do not have the same direction.

Table 4a. Distribution of Health Status by Household Socioeconomic Status in Childhood

Sample of all individuals

a. Quintile 5 vs. Quintile 1

Health Status	Quintile 1			Quintile 5			$\widehat{\Delta F}(k)$	z_k^l
	Freq.	%	Cumul. %	Freq.	%	Cumul. %		
1 = Poor/Fair	300	35.5	35.5	43	11.3	11.3	-0.243	-0.243 ***
2 = Good	508	60.1	95.6	258	68.2	79.5	-0.161	-0.161 ***
3 = Excellent	37	4.4	100.0	78	20.5	100.0		
Total	845	100.0		378	100.0			

b. Quintile 5 vs. Quintile 2

Health Status	Quintile 2			Quintile 5			$\widehat{\Delta F}(k)$	z_k^l
	Freq.	%	Cumul. %	Freq.	%	Cumul. %		
1 = Poor/Fair	91	26.7	26.7	43	11.3	11.3	-0.154	-5.319 ***
2 = Good	227	66.9	93.5	258	68.2	79.5	-0.140	-5.683 ***
3 = Excellent	22	6.5	100.0	78	20.5	100.0		
Total	340	100.0		378	100.0			

c. Quintile 5 vs. Quintile 3

Health Status	Quintile 3			Quintile 5			$\widehat{\Delta F}(k)$	z_k^l
	Freq.	%	Cumul. %	Freq.	%	Cumul. %		
1 = Poor/Fair	45	17.6	17.6	43	11.3	11.3	-0.063	-2.194 ***
2 = Good	199	78.2	95.8	258	68.2	79.5	-0.163	-6.720 ***
3 = Excellent	11	4.2	100.0	78	20.5	100.0		
Total	255	100.0		378	100.0			

d. Quintile 5 vs. Quintile 4

Health Status	Quintile 4			Quintile 5			$\widehat{\Delta F}(k)$	z_k^l
	Freq.	%	Cumul. %	Freq.	%	Cumul. %		
1 = Poor/Fair	79	20.0	20.0	43	11.3	11.3	-0.087	-3.375 ***
2 = Good	278	70.2	90.1	258	68.2	79.5	-0.106	-4.150 ***
3 = Excellent	39	9.9	100.0	78	20.5	100.0		
Total	396	100.0		378	100.0			

e. Quintile 4 vs. Quintile 1

Health Status	Quintile 1			Quintile 4			$\widehat{\Delta F}(k)$	z_k^l
	Freq.	%	Cumul. %	Freq.	%	Cumul. %		
1 = Poor/Fair	300	35.5	35.5	79	19.97	19.97	-0.155	-5.983 ***
2 = Good	508	60.1	95.6	278	70.15	90.12	-0.055	-3.300 ***
3 = Excellent	37	4.4	100.0	39	9.88	100		
Total	845	100.0		396	100			

f. Quintile 4 vs. Quintile 2

Health Status	Quintile 2			Quintile 4			$\widehat{\Delta F}(k)$	z_k^l
	Freq.	%	Cumul. %	Freq.	%	Cumul. %		
1 = Poor/Fair	91	26.7	26.7	79	20.0	20.0	-0.067	-2.139 ***
2 = Good	227	66.9	93.5	278	70.2	90.1	-0.034	-1.693 ***
3 = Excellent	22	6.5	100.0	39	9.9	100.0		
Total	340	100.0		396	100.0			

g. Quintile 4 vs. Quintile 3

Health Status	Quintile 3			Quintile 4			$\widehat{\Delta F}(k)$	z_k^l
	Freq.	%	Cumul. %	Freq.	%	Cumul. %		
1 = Poor/Fair	45	17.6	17.6	70	20.0	20.0	0.024	0.767
2 = Good	199	78.2	95.8	310	70.2	90.1	-0.057	-2.904 ***
3 = Excellent	11	4.2	100.0	17	9.9	100.0		
Total	255	100.0		396	100.0			

h. Quintile 3 vs. Quintile 1

Health Status	Quintile 1			Quintile 3			$\widehat{\Delta F}(k)$	z_k^l
	Freq.	%	Cumul. %	Freq.	%	Cumul. %		
1 = Poor/Fair	300	35.5	35.5	45	17.6	17.6	-0.179	-0.293
2 = Good	508	60.1	95.6	199	78.2	95.8	0.002	0.007
3 = Excellent	37	4.4	100.0	11	4.2	100.0		
Total	845	100.0		255	100.0			

i. Quintile 3 vs. Quintile 2

Health Status	Quintile 2			Quintile 3			$\widehat{\Delta F}(k)$	z_k^l
	Freq.	%	Cumul. %	Freq.	%	Cumul. %		
1 = Poor/Fair	91	26.7	26.7	45	17.6	17.6	-0.091	-2.685 ***
2 = Good	227	66.9	93.5	199	78.2	95.8	0.023	1.244
3 = Excellent	22	6.5	100.0	11	4.2	100.0		
Total	340	100.0		255	100.0			

j. Quintile 2 vs. Quintile 1

Health Status	Quintile 1			Quntile 2			$\widehat{\Delta F}(k)$	z_k^l
	Freq.	%	Cumul. %	Freq.	%	Cumul. %		
1 = Poor/Fair	300	35.5	35.5	91	26.7	26.7	-0.089	-3.043 ***
2 = Good	508	60.1	95.6	227	66.9	93.5	-0.021	-1.371
3 = Excellent	37	4.4	100.0	22	6.5	100.0		
Total	845	100.0		340	100.0			

*** denote that the statistic is significant at the 5% significant level

Source: 2010 Colombian LSSM Survey

Note: The null hypothesis is given by $H_0 = \Delta F(k) \geq 0$ for some $k \in \{1,2\}$ and the alternative is given by $H_1 = \Delta F(k) < 0$ for all $k \in \{1,2\}$. $\widehat{\Delta F}(k)$ indicates the estimated difference between the cumulative probability functions, $\widehat{F_B(k)}$–$\widehat{F_A(k)}$, where $\widehat{F_B(k)}$ indicates the cumulative probability function for the subgroup in the most-right panel and $\widehat{F_A(k)}$ for the most-left panel, for row k. H_0 is rejected if and only if $z_k^l \leq -z^* < 0$ for all $k \in \{1,2\}$, where $-z^* = -1.645$ is the left-tail critical value at the 5% significance level. No ordering can be established if the two values for z_k^l do not have the same direction.

Table 4b. Distribution of Health Status by Household Socioeconomic Status in Childhood: Residents in Urban Areas

a. Quintile 5 vs. Quintile 1

Health Status	Quintile 1			Quintile 5			$\widehat{\Delta F}(k)$	z_k^l
	Freq.	%	Cumul. %	Freq.	%	Cumul. %		
1 = Poor/Fair	103	33.5	33.5	36	10.9	10.9	-0.226	-7.086 ***
2 = Good	189	61.2	94.7	222	67.8	78.7	-0.160	-6.173 ***
3 = Excellent	16	5.3	100.0	70	21.3	100.0		
Total	308	100.0		327	100.0			

b. Quintile 5 vs. Quintile 2

Health Status	Quintile 2			Quintile 5			$\widehat{\Delta F}(k)$	z_k^l
	Freq.	%	Cumul. %	Freq.	%	Cumul. %		
1 = Poor/Fair	50	25.3	25.3	36	10.9	10.9	-0.144	-4.065 ***
2 = Good	135	67.9	93.2	222	67.8	78.7	-0.145	-5.012 ***
3 = Excellent	14	6.8	100.0	70	21.3	100.0		
Total	198	100.0		327	100.0			

c. Quintile 5 vs. Quintile 3

Health Status	Quintile 3			Quintile 5			$\widehat{\Delta F}(k)$	z_k^l
	Freq.	%	Cumul. %	Freq.	%	Cumul. %		
1 = Poor/Fair	24	16.2	16.2	36	10.9	10.9	-0.053	-1.528
2 = Good	118	79.1	95.3	222	67.8	78.7	-0.166	-5.801 ***
3 = Excellent	7	4.7	100.0	70	21.3	100.0		
Total	149	100.0		327	100.0			

d. Quintile 5 vs. Quintile 4

Health Status	Quintile 4			Quintile 5			$\widehat{\Delta F}(k)$	z_k^l
	Freq.	%	Cumul. %	Freq.	%	Cumul. %		
1 = Poor/Fair	55	20.6	20.6	36	10.9	10.9	-0.097	-3.219 ***
2 = Good	185	69.1	89.7	222	67.8	78.7	-0.110	-3.746 ***
3 = Excellent	28	10.3	100.0	70	21.3	100.0		
Total	268	100.0		327	100.0			

e. Quintile 4 vs. Quintile 1

Health Status	Quintile 1			Quintile 4			$\widehat{\Delta F}(k)$	z_k^l
	Freq.	%	Cumul. %	Freq.	%	Cumul. %		
1 = Poor/Fair	103	33.5	33.5	55	20.57	20.57	-0.129	-3.544 ***
2 = Good	189	61.2	94.7	185	69.11	89.68	-0.051	-2.247 ***
3 = Excellent	16	5.3	100.0	28	10.32	100		
Total	308	100.0		268	100			

f. Quintile 4 vs. Quintile 2

Health Status	Quintile 2			Quintile 4			$\widehat{\Delta F}(k)$	z_k^l
	Freq.	%	Cumul. %	Freq.	%	Cumul. %		
1 = Poor/Fair	50	25.3	25.3	55	20.6	20.6	-0.047	-1.184
2 = Good	135	67.9	93.2	185	69.1	89.7	-0.035	-1.356
3 = Excellent	14	6.8	100.0	28	10.3	100.0		
Total	198	100.0		268	100.0			

g. Quintile 4 vs. Quintile 3

Health Status	Quintile 3			Quintile 4			$\widehat{\Delta F}(k)$	z_k^l
	Freq.	%	Cumul. %	Freq.	%	Cumul. %		
1 = Poor/Fair	24	16.2	16.2	43	20.6	20.6	0.044	1.123
2 = Good	118	79.1	95.3	212	69.1	89.7	-0.056	-2.196 ***
3 = Excellent	7	4.7	100.0	13	10.3	100.0		
Total	149	100.0		268	100.0			

h. Quintile 3 vs. Quintile 1

Health Status	Quintile 1			Quintile 3			$\widehat{\Delta F}(k)$	z_k^l
	Freq.	%	Cumul. %	Freq.	%	Cumul. %		
1 = Poor/Fair	103	33.5	33.5	24	16.2	16.2	-0.173	-0.289
2 = Good	189	61.2	94.7	118	79.1	95.3	0.005	0.017
3 = Excellent	16	5.3	100.0	7	4.7	100.0		
Total	308	100.0		149	100.0			

i. Quintile 3 vs. Quintile 2

Health Status	Quintile 2			Quintile 3			$\widehat{\Delta F}(k)$	z_k^l
	Freq.	%	Cumul. %	Freq.	%	Cumul. %		
1 = Poor/Fair	50	25.3	25.3	24	16.2	16.2	-0.091	-2.099 ***
2 = Good	135	67.9	93.2	118	79.1	95.3	0.021	0.837
3 = Excellent	14	6.8	100.0	7	4.7	100.0		
Total	198	100.0		149	100.0			

j. Quintile 2 vs. Quintile 1

Health Status	Quintile 1			Quntile 2			$\widehat{\Delta F}(k)$	z_k^l
	Freq.	%	Cumul. %	Freq.	%	Cumul. %		
1 = Poor/Fair	103	33.5	33.5	50	25.3	25.3	-0.083	-2.017 ***
2 = Good	189	61.2	94.7	135	67.9	93.2	-0.016	-0.710
3 = Excellent	16	5.3	100.0	14	6.8	100.0		
Total	308	100.0		198	100.0			

*** denote that the statistic is significant at the 5% significant level

Source: 2010 Colombian LSSM Survey

Note: The null hypothesis is given by $H_0 = \Delta F(k) \geq 0$ for some $k \in \{1,2\}$ and the alternative is given by $H_1 = \Delta F(k) < 0$ for all $k \in \{1,2\}$. $\widehat{\Delta F}(k)$ indicates the estimated difference between the cumulative probability functions, $\widehat{F_B(k)} - \widehat{F_A(k)}$, where $\widehat{F_B(k)}$ indicates the cumulative probability function for the subgroup in the most-right panel and $\widehat{F_A(k)}$ for the most-left panel, for row k. H_0 is rejected if and only if $z_k^l \leq -z^* < 0$ for all $k \in \{1,2\}$, where $-z^* = -1.645$ is the left-tail critical value at the 5% significance level. No ordering can be established if the two values for z_k^l do not have the same direction.

Table 4c. Distribution of Health Status by Household Socioeconomic Status in Childhood: Residents in Rural Areas

a. Quintile 5 vs. Quintile 1

Health Status	Quintile 1			Quintile 5			$\widehat{\Delta F}(k)$	z_k^l
	Freq.	%	Cumul. %	Freq.	%	Cumul. %		
1 = Poor/Fair	209	38.8	38.8	9	18.4	18.4	-0.204	-3.508 ***
2 = Good	312	58.2	97.0	39	76.1	94.5	-0.025	-0.761
3 = Excellent	16	3.0	100.0	3	5.5	100.0		
Total	537	100.0		51	100.0			

b. Quintile 5 vs. Quintile 2

Quintile 2			Quintile 5			$\widehat{\Delta F}(k)$	z_k^l
Freq.	%	Cumul. %	Freq.	%	Cumul. %		
46	32.3	32.3	9	18.4	18.4	-0.139	-2.070 ***
89	62.6	94.8	39	76.1	94.5	-0.003	-0.087
7	5.2	100.0	3	5.5	100.0		
142	100.0		51	100.0			

c. Quintile 5 vs. Quintile 3

Health Status	Quintile 3			Quintile 5			$\widehat{\Delta F}(k)$	z_k^l
	Freq.	%	Cumul. %	Freq.	%	Cumul. %		
1 = Poor/Fair	25	23.1	23.1	9	18.4	18.4	-0.047	-0.693
2 = Good	79	74.8	97.9	39	76.1	94.5	-0.034	-0.972
3 = Excellent	2	2.1	100.0	3	5.5	100.0		
Total	106	100.0		51	100.0			

d. Quintile 5 vs. Quintile 4

Quintile 4			Quintile 5			$\widehat{\Delta F}(k)$	z_k^l
Freq.	%	Cumul. %	Freq.	%	Cumul. %		
21	16.3	16.3	9	18.4	18.4	0.021	0.328
98	76.4	92.7	39	76.1	94.5	0.018	0.461
9	7.3	100.0	3	5.5	100.0		
128	100.0		51	100.0			

e. Quintile 4 vs. Quintile 1

Health Status	Quintile 1			Quintile 4			$\widehat{\Delta F}(k)$	z_k^l
	Freq.	%	Cumul. %	Freq.	%	Cumul. %		
1 = Poor/Fair	209	38.8	38.8	21	16.33	16.33	-0.225	-5.791 ***
2 = Good	312	58.2	97.0	98	76.38	92.71	-0.043	-1.782 ***
3 = Excellent	16	3.0	100.0	9	7.29	100		
Total	537	100.0		128	100			

f. Quintile 4 vs. Quintile 2

Quintile 2			Quintile 4			$\widehat{\Delta F}(k)$	z_k^l
Freq.	%	Cumul. %	Freq.	%	Cumul. %		
46	32.3	32.3	21	16.3	16.3	-0.159	-3.122 ***
89	62.6	94.8	98	76.4	92.7	-0.021	-0.721
7	5.2	100.0	9	7.3	100.0		
142	100.0		128	100.0			

g. Quintile 4 vs. Quintile 3

Health Status	Quintile 3			Quintile 4			$\widehat{\Delta F}(k)$	z_k^l
	Freq.	%	Cumul. %	Freq.	%	Cumul. %		
1 = Poor/Fair	25	23.1	23.1	30	16.3	16.3	-0.068	-1.296
2 = Good	79	74.8	97.9	96	76.4	92.7	-0.052	-1.932 ***
3 = Excellent	2	2.1	100.0	3	7.3	100.0		
Total	106	100.0		128	100.0			

h. Quintile 3 vs. Quintile 1

Quintile 1			Quintile 3			$\widehat{\Delta F}(k)$	z_k^l
Freq.	%	Cumul. %	Freq.	%	Cumul. %		
209	38.8	38.8	25	23.1	23.1	-0.157	-0.244
312	58.2	97.0	79	74.8	97.9	0.009	0.040
16	3.0	100.0	2	2.1	100.0		
537	100.0		106	100.0			

i. Quintile 3 vs. Quintile 2

Health Status	Quintile 2			Quintile 3			$\widehat{\Delta F}(k)$	z_k^l
	Freq.	%	Cumul. %	Freq.	%	Cumul. %		
1 = Poor/Fair	46	32.3	32.3	25	23.1	23.1	-0.092	-1.613
2 = Good	89	62.6	94.8	79	74.8	97.9	0.031	1.319
3 = Excellent	7	5.2	100.0	2	2.1	100.0		
Total	142	100.0		106	100.0			

j. Quintile 2 vs. Quintile 1

Quintile 1			Quntile 2			$\widehat{\Delta F}(k)$	z_k^l
Freq.	%	Cumul. %	Freq.	%	Cumul. %		
209	38.8	38.8	46	32.3	32.3	-0.066	-1.474
312	58.2	97.0	89	62.6	94.8	-0.022	-1.087
16	3.0	100.0	7	5.2	100.0		
537	100.0		142	100.0			

*** denote that the statistic is significant at the 5% significant level

Source: 2010 Colombian LSSM Survey

Note: The null hypothesis is given by $H_0 = \Delta F(k) \geq 0$ for some $k \in \{1,2\}$ and the alternative is given by $H_1 = \Delta F(k) < 0$ for all $k \in \{1,2\}$. $\widehat{\Delta F}(k)$ indicates the estimated difference between the cumulative probability functions, $\widehat{F_B(k)}$–$\widehat{F_A(k)}$, where $\widehat{F_B(k)}$ indicates the cumulative probability function for the subgroup in the most-right panel and $\widehat{F_A(k)}$ for the most-left panel, for row k. H_0 is rejected if and only if $z_k^l \leq -z^* < 0$ for all $k \in \{1,2\}$, where $-z^* = -1.645$ is the left-tail critical value at the 5% significance level. No ordering can be established if the two values for z_k^l do not have the same direction.

Table 5. Distribution of Health Status by Parental Vital Status

All individuals

Health Status	Father Alive			Father Deceased			$\widehat{\Delta F}(k)$	z_k^l
	Freq.	%	Cumul. %	Freq.	%	Cumul. %		
1 = Poor/Fair	199	18.57	18.57	332	28.38	28.38	-0.098	-5.527 ***
2 = Good	736	68.76	87.33	766	65.54	93.92	-0.066	-5.342 ***
3 = Excellent	136	12.67	100	71	6.08	100		
Total	1,071	100		1,169	100			

Health Status	Mother Alive			Mother Deceased			$\widehat{\Delta F}(k)$	z_k^l
	Freq.	%	Cumul. %	Freq.	%	Cumul. %		
1 = Poor/Fair	286	19.61	19.61	246	31.46	31.46	-0.119	-6.046 ***
2 = Good	1027	70.36	89.97	474	60.63	92.09	-0.021	-1.702 ***
3 = Excellent	146	10.03	100	62	7.92	100.01		
Total	1,459	100		781	100.01			

Residents in Urban Areas

Health Status	Father Alive			Father Deceased			$\widehat{\Delta F}(k)$	z_k^l
	Freq.	%	Cumul. %	Freq.	%	Cumul. %		
1 = Poor/Fair	98	16.28	16.28	172	26.23	26.23	-0.100	-4.356 ***
2 = Good	412	68.57	84.85	441	67.16	93.39	-0.085	-4.866 ***
3 = Excellent	91	15.14	99.99	43	6.61	100		
Total	601	99.99		656	100			

Health Status	Mother Alive			Mother Deceased			$\widehat{\Delta F}(k)$	z_k^l
	Freq.	%	Cumul. %	Freq.	%	Cumul. %		
1 = Poor/Fair	145	17.45	17.45	125	29.34	29.34	-0.119	-4.628 ***
2 = Good	589	70.87	88.32	263	61.68	91.02	-0.027	-1.519
3 = Excellent	97	11.68	100	38	8.97	99.99		
Total	831	100		426	99.99			

Residents in Rural Areas

Health Status	Father Alive			Father Deceased			$\widehat{\Delta F}(k)$	z_k^l
	Freq.	%	Cumul. %	Freq.	%	Cumul. %		
1 = Poor/Fair	126	26.82	26.82	189	36.77	36.77	-0.100	-3.372 ***
2 = Good	326	69.45	96.27	304	59.23	96	0.003	0.220
3 = Excellent	18	3.73	100	21	4	100		
Total	470	100		513	100			

Health Status	Mother Alive			Mother Deceased			$\widehat{\Delta F}(k)$	z_k^l
	Freq.	%	Cumul. %	Freq.	%	Cumul. %		
1 = Poor/Fair	175	27.87	27.87	139	39.09	39.09	-0.112	-3.565 ***
2 = Good	429	68.39	96.26	202	56.81	95.9	0.004	0.278
3 = Excellent	23	3.74	100	15	4.11	100.01		
Total	628	100		355	100.01			

*** denote that the statistic is significant at the 5% significant level
Source: 2010 Colombian LSSM Survey

Note: The null hypothesis is given by $H_0 = \Delta F(k) \geq 0$ for some $k \in \{1,2\}$ and the alternative is given by $H_1 = \Delta F(k) < 0$ for all $k \in \{1,2\}$. $\widehat{\Delta F}(k)$ indicates the estimated difference between the cumulative probability functions, $\widehat{F_B(k)}$–$\widehat{F_A(k)}$, where $\widehat{F_B(k)}$ indicates the cumulative probability function for the subgroup in the most-right panel and $\widehat{F_A(k)}$ for the most-left panel, for row k. H_0 is rejected if and only if $z_k^l \leq -z^* < 0$ for all $k \in \{1,2\}$, where $-z^* = -1.645$ is the left-tail critical value at the 5% significance level. No ordering can be established if the two values for z_k^l do not have the same direction.

Table 6. Log-odds ratios for the Correlates of Health Status

	All Individuals			Urban Areas			Rural Areas		
	Logistic regression for self-assessed health		OLS for years of education	Logistic regression for self-assessed health		OLS for years of education	Logistic regression for self-assessed health		OLS for years of education
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
Male	0.5718***	0.5888***	0.1485	0.6480***	0.7090***	0.5516**	0.5238**	0.4682**	-0.5169*
	(0.1308)	(0.1309)	(0.1971)	(0.1615)	(0.1615)	(0.2307)	(0.2070)	(0.2057)	(0.2763)
Age group:									
35-45 years old	-0.5438***	-0.5265***	0.1513	-0.5249*	-0.5017*	0.2096	-0.5885**	-0.6504***	-0.5751*
	(0.2030)	(0.2030)	(0.2310)	(0.2794)	(0.2794)	(0.2785)	(0.2524)	(0.2523)	(0.3100)
45-55 years old	-0.7722***	-0.7712***	0.0091	-0.7791***	-0.7919***	-0.1165	-0.9601***	-1.0357***	-0.7022**
	(0.2035)	(0.2035)	(0.2567)	(0.2745)	(0.2743)	(0.3035)	(0.2608)	(0.2622)	(0.3539)
55-65 years old	-1.3343***	-1.4179***	-0.7311**	-1.3810***	-1.4865***	-0.9536***	-1.5138***	-1.6726***	-1.4749***
	(0.2210)	(0.2211)	(0.3002)	(0.2945)	(0.2946)	(0.3650)	(0.2895)	(0.2913)	(0.3759)
Ethnicity:									
Indigenous	-0.1846	-0.2064	-0.1902	-0.6067	-0.6089	-0.0198	0.3515	0.3396	-0.1108
	(0.4418)	(0.4416)	(0.5689)	(0.6196)	(0.6196)	(0.7708)	(0.4447)	(0.4447)	(0.6348)
Black/mulato/raizal/palenquero	-0.2954	-0.2842	0.0977	-0.4191	-0.4389	-0.1797	0.1126	0.1441	0.2919
	(0.2374)	(0.2373)	(0.3620)	(0.2967)	(0.2968)	(0.4474)	(0.3392)	(0.3389)	(0.4049)
Born in urban area	-0.0626	0.0900	1.3345***	-0.2093	-0.1340	0.6813**	0.1539	0.1819	0.2604
	(0.1382)	(0.1372)	(0.2150)	(0.1819)	(0.1818)	(0.2760)	(0.2400)	(0.2399)	(0.2870)
Household socioeconomic status at age 10:									
Quintile 2	0.2254	0.3610**	1.1864***	0.2705	0.3896*	1.0771***	0.0221	0.1019	0.7404**
	(0.1728)	(0.1726)	(0.2786)	(0.2260)	(0.2265)	(0.3511)	(0.2306)	(0.2286)	(0.3251)
Quintile 3	0.4432**	0.6787***	2.0605***	0.4838*	0.7090**	2.0365***	0.3046	0.4438	1.2923***
	(0.2192)	(0.2190)	(0.3221)	(0.2856)	(0.2863)	(0.4052)	(0.2942)	(0.2897)	(0.3834)
Quintile 4	-0.0220	0.2894	2.7243***	-0.1321	0.1571	2.6149***	0.6039**	0.8141***	1.9519***
	(0.1924)	(0.1866)	(0.2949)	(0.2408)	(0.2323)	(0.3556)	(0.3056)	(0.2955)	(0.3979)
Quintile 5	0.3424	0.7875***	3.8949***	0.3067	0.7039***	3.5914***	0.1654	0.5132	3.2298***
	(0.2430)	(0.2330)	(0.3100)	(0.2817)	(0.2697)	(0.3618)	(0.4652)	(0.4507)	(0.5655)
Paternal education level:									
Complete primary and incomplete secondary	0.3279	0.4472**	1.0441***	0.4544*	0.5522**	0.8835**	-0.3063	-0.1908	1.0720**
	(0.2210)	(0.2208)	(0.3074)	(0.2602)	(0.2604)	(0.3561)	(0.3666)	(0.3642)	(0.5066)
Complete secondary or more	-0.0880	0.1332	1.9346***	-0.0830	0.0965	1.6231***	0.4943	0.8972	3.7419***
	(0.3775)	(0.3790)	(0.4178)	(0.4071)	(0.4088)	(0.4658)	(0.8467)	(0.8426)	(0.7677)
Unknown father's level of education	0.1154	0.0374	-0.6816**	0.3351	0.2557	-0.7175*	-0.3142	-0.3670	-0.4904*
	(0.1973)	(0.1971)	(0.2847)	(0.2712)	(0.2709)	(0.3668)	(0.2422)	(0.2407)	(0.2933)
Maternal education level:									
Complete primary and incomplete secondary	-0.0310	0.0893	1.0529***	0.0025	0.1204	1.0656***	-0.2464	-0.1770	0.6437
	(0.2113)	(0.2109)	(0.2937)	(0.2559)	(0.2553)	(0.3440)	(0.3102)	(0.3094)	(0.4030)
Complete secondary or more	0.6535	0.9660**	2.7346***	0.9304*	1.2408**	2.8059***	-1.0822	-0.8596	2.0676**
	(0.4436)	(0.4387)	(0.4138)	(0.5102)	(0.5047)	(0.4616)	(0.7439)	(0.7402)	(0.9935)
Unknown mother's level of education	-0.0639	-0.1039	-0.3505	-0.1101	-0.1278	-0.1596	0.0380	0.0143	-0.2207
	(0.2326)	(0.2324)	(0.3570)	(0.3235)	(0.3235)	(0.4762)	(0.2686)	(0.2687)	(0.3324)
Parental health-status:									
Father alive	0.0271	0.0173	-0.0855	0.0651	0.0576	-0.0681	-0.1580	-0.1770	-0.1761
	(0.1385)	(0.1384)	(0.2071)	(0.1747)	(0.1746)	(0.2468)	(0.1887)	(0.1887)	(0.2698)
Mother alive	-0.0747	-0.0211	0.4691**	-0.0905	-0.0577	0.2965	-0.0151	0.0630	0.7250***
	(0.1466)	(0.1464)	(0.2299)	(0.1844)	(0.1844)	(0.2761)	(0.2058)	(0.2048)	(0.2730)
Years of education	0.1143***			0.1106***			0.1077***		
	(0.0171)			(0.0212)			(0.0267)		
Years of education purged from circumstances		0.1143***			0.1106***			0.1077***	
		(0.0171)			(0.0212)			(0.0267)	
Constant	0.7065**	1.1524***	3.9011***	0.7212*	1.2670***	4.9344***	1.1232***	1.5748***	4.1936***
	(0.2803)	(0.2739)	(0.4113)	(0.3910)	(0.3772)	(0.5439)	(0.3710)	(0.3678)	(0.5162)
Observations	2,203	2,203	2,205	1,239	1,239	1,241	957	957	958
Region of Birth Dummy	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Log-likelihood	-4.460e+06	-4.460e+06		-3.304e+06	-3.304e+06		-1.096e+06	-1.096e+06	
Pseudo R squared	0.128	0.128		0.140	0.140		0.104	0.104	
Adjusted R squared			0.413			0.377			0.219

***, **, and * indicate statistical significance at the 1, 5 and 10 percent level, respectively.
Robust standard errors in parentheses
Own calculations. Source: 2010 Colombian LSSM

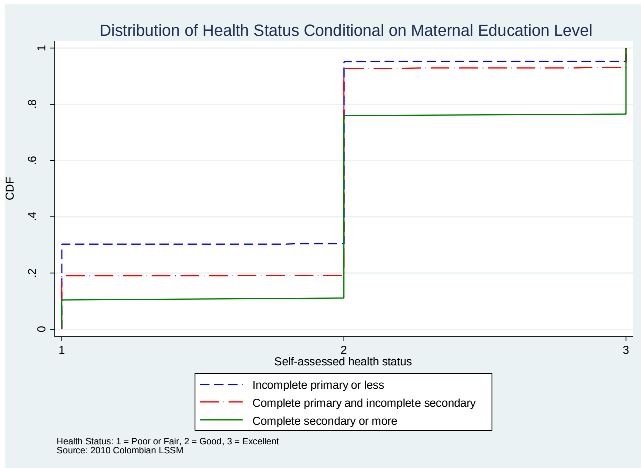
Table 7. Decomposition of the Dissimilarity Index of Inequality of Opportunity

	All individuals		Residents in Urban Areas		Residents in Rural Areas	
Dissimilarity Index	0.0847	0.0847	0.0802	0.0803	0.1014	0.1012
	Decomposition of the Dissimilarity Index (in %)					
Educational Attainment	46.635		42.687		38.732	
Education purged from circumstances		34.110		32.789		31.999
Circumstances, of which:	53.365	65.890	57.313	67.212	61.268	68.001
Ethnicity	1.528	1.577	3.902	4.025	0.920	0.953
Region of Birth	4.270	5.402	11.689	11.804	24.597	24.369
Born in Urban Area	10.464	10.708	1.071	0.944	4.218	4.937
Household Socioeconomic Status at age 10	14.870	20.448	12.840	16.993	19.218	25.416
Mother's Education	10.460	13.626	13.687	17.480	4.115	2.551
Father's Education	10.338	12.516	12.679	14.312	7.204	9.130
Mother's Vital Status	1.223	1.399	1.335	1.540	0.995	0.595
Father's Vital Status	0.213	0.214	0.109	0.113	0.001	0.051
Observations	2,203		1,239		957	

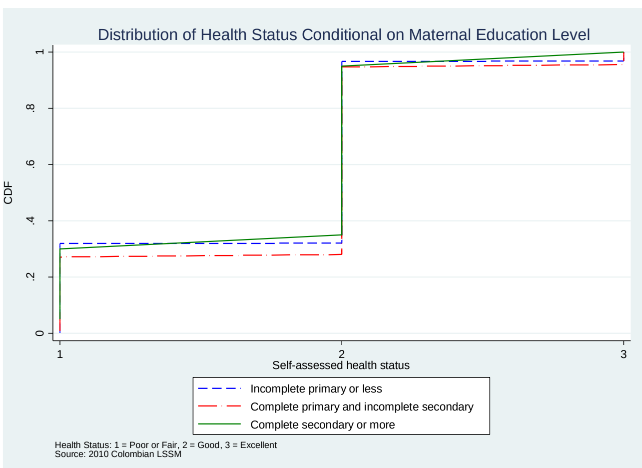
Own calculations. Source: 2010 Colombian LSSM

Figure 1. Dominance according to Mother's Education Level

Sample of all individuals



Rural sample



Urban sample

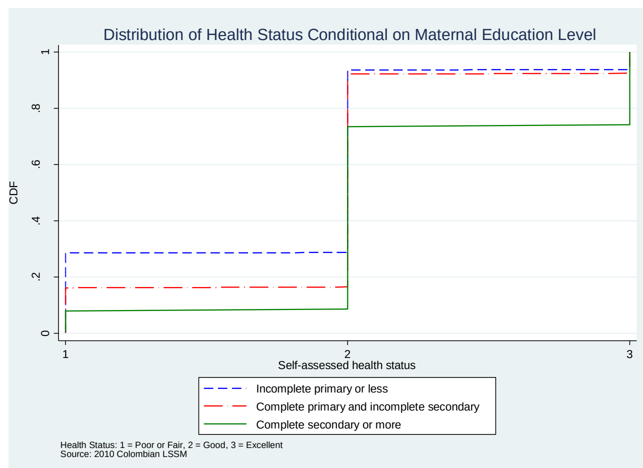
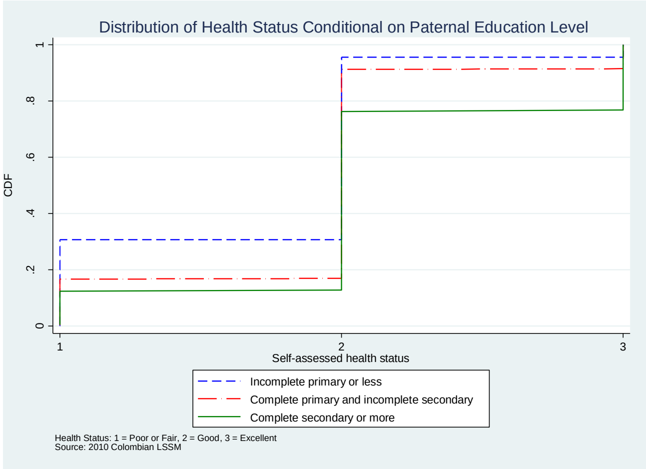
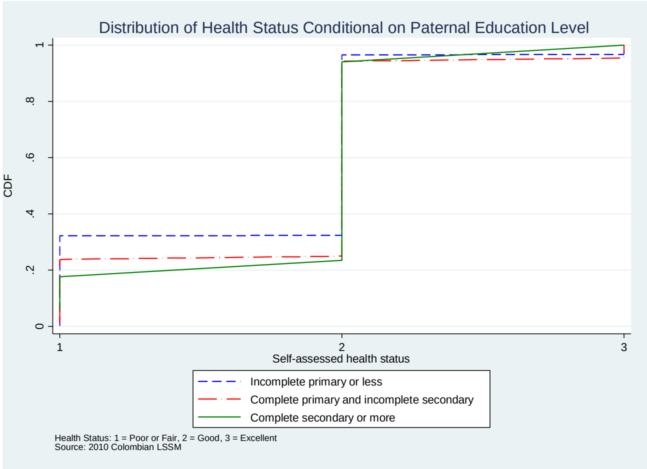


Figure 2. Dominance according to Father's Education Level

Sample of all individuals



Rural sample



Urban sample

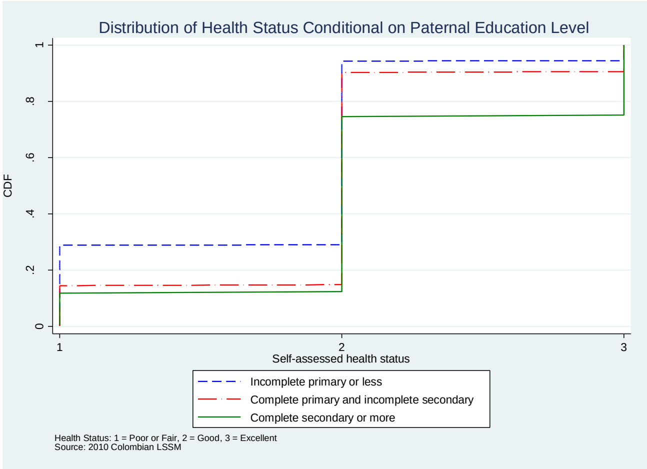
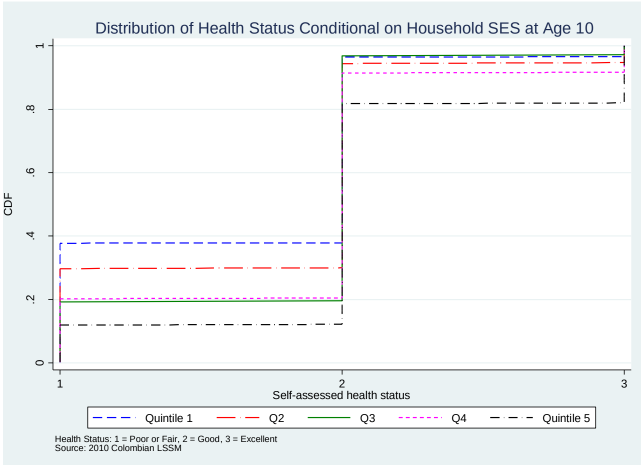
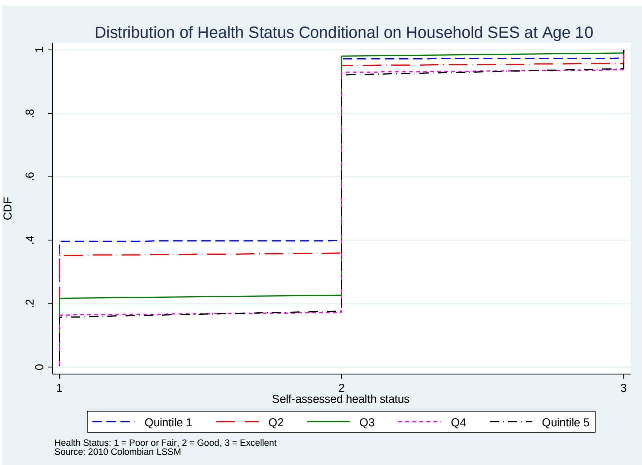


Figure 3. Dominance according to Household Socioeconomic Status at age 10

Sample of all individuals



Rural sample



Urban sample

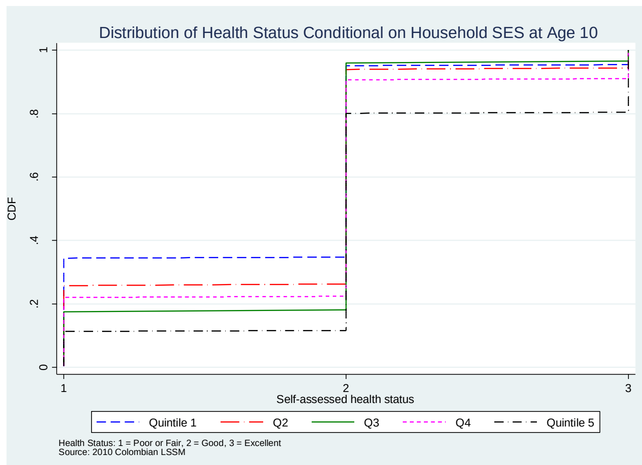
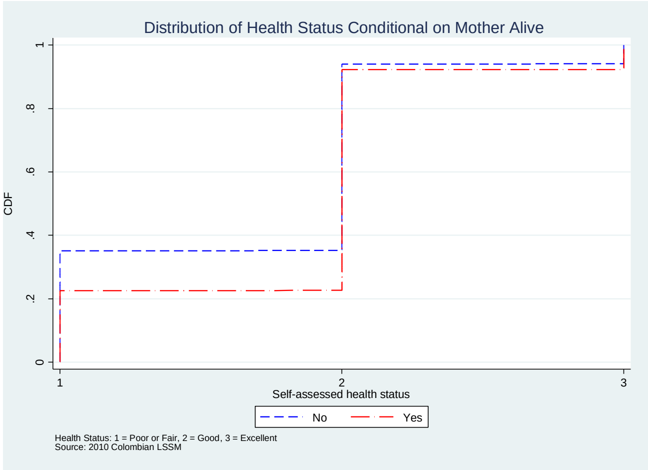
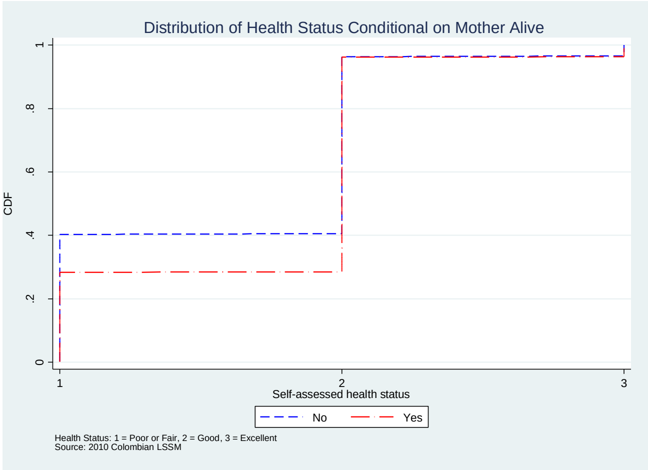


Figure 4. Dominance according to Mother's Vital Status

Sample of all individuals



Rural sample



Urban sample

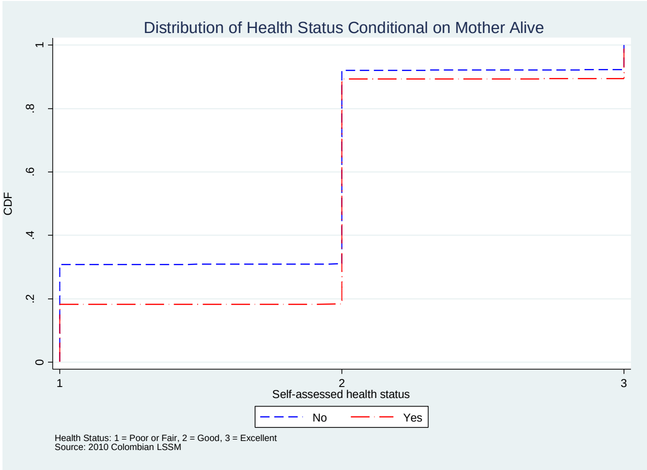
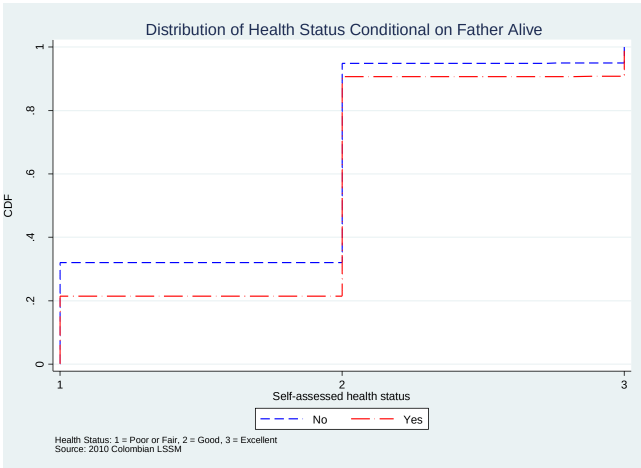
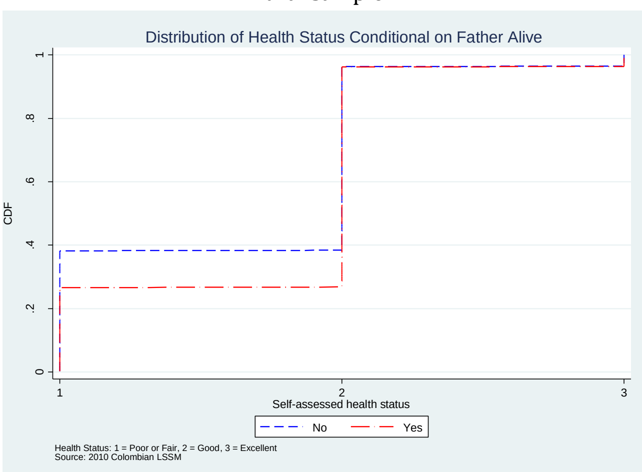


Figure 5. Dominance according to Father's Vital Status

Sample of all individuals



Rural sample



Urban sample